

Declaration of Konstantinos Psounis

Unredacted Version of Document

QUINN EMANUEL URQUHART & SULLIVAN, LLP

Diane M. Doolittle (CA Bar No. 142046)
dianedoolittle@quinnemanuel.com
Sara Jenkins (CA Bar No. 230097)
sarajenkins@quinnemanuel.com
555 Twin Dolphin Drive, 5th Floor
Redwood Shores, CA 94065
Telephone: (650) 801-5000
Facsimile: (650) 801-5100

Andrew H. Schapiro (admitted *pro hac vice*)
andrewschapiro@quinnemanuel.com
Teuta Fani (admitted *pro hac vice*)
teutafani@quinnemanuel.com
191 N. Wacker Drive, Suite 2700
Chicago, IL 60606
Telephone: (312) 705-7400
Facsimile: (312) 705-7401

Stephen A. Broome (CA Bar No. 314605)
stephenbroome@quinnemanuel.com
Viola Trebicka (CA Bar No. 269526)
violatrebicka@quinnemanuel.com
Crystal Nix-Hines (Bar No. 326971)
crystalnixhines@quinnemanuel.com
Alyssa G. Olson (CA Bar No. 305705)
alyolson@quinnemanuel.com
865 S. Figueroa Street, 10th Floor
Los Angeles, CA 90017
Telephone: (213) 443-3000
Facsimile: (213) 443-3100

Josef Ansorge (admitted *pro hac vice*)
josefansorge@quinnemanuel.com
Xi ("Tracy") Gao (CA Bar No. 326266)
tracygao@quinnemanuel.com
Carl Spilly (admitted *pro hac vice*)
carlspilly@quinnemanuel.com
1300 I Street NW, Suite 900
Washington D.C., 20005
Telephone: (202) 538-8000
Facsimile: (202) 538-8100

Jomaire Crawford (admitted *pro hac vice*)
jomairecrawford@quinnemanuel.com
51 Madison Avenue, 22nd Floor
New York, NY 10010
Telephone: (212) 849-7000
Facsimile: (212) 849-7100

Jonathan Tse (CA Bar No. 305468)
jonathantse@quinnemanuel.com
50 California Street, 22nd Floor
San Francisco, CA 94111
Telephone: (415) 875-6600
Facsimile: (415) 875-6700

Attorneys for Defendant Google LLC

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA, OAKLAND DIVISION

CHASOM BROWN, WILLIAM BYATT,
JEREMY DAVIS, CHRISTOPHER
CASTILLO, and MONIQUE TRUJILLO,
individually and on behalf of all similarly
situated,

Plaintiffs,

v.

GOOGLE LLC,
Defendant.

Case No. 4:20-cv-03664-YGR-SVK

**DECLARATION OF KONSTANTINOS
PSOUNIS IN SUPPORT OF GOOGLE,
LLC'S OPPOSITION TO PLAINTIFF'S
MOTION FOR CLASS CERTIFICATION
AND APPOINTMENT OF CLASS
REPRESENTATIVES AND CLASS
COUNSEL**

Judge: Hon. Yvonne Gonzalez Rogers
Hearing Date: September 20, 2022
Hearing Time: 2:00 p.m..

1 I, Konstantinos Psounis, declare as follows:

2 1. Counsel for Defendant Google, LLC retained me to provide expert analysis and, if
3 requested, expert testimony in this matter.

4 2. I submit this declaration in support of Google's Opposition to Plaintiff's Motion for
5 Class Certification.

6 3. Attached as Exhibit 1 is a true and correct copy of the Expert Report of Konstantinos
7 Psounis, dated June 7, 2022. The opinions I provided therein are true and correct to the best of my
8 knowledge.

9
10 I declare under penalty of perjury of the laws of the United States that the foregoing is true
11 and correct. Executed in Athens, Greece on July 29, 2022

12
13 By 
14 Konstantinos Psounis
15
16
17
18
19
20
21
22
23
24
25
26
27
28

EXHIBIT 1

**Redacted Version of
Document Sought to
be Sealed**

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA – OAKLAND DIVISION

CHASOM BROWN, WILLIAM BYATT,
JEREMY DAVIS, CHRISTOPHER
CASTILLO, and MONIQUE TRUJILLO,
individually and on behalf of all similarly
situated,

Plaintiffs,

v.

GOOGLE LLC,

Defendant.

Case No. 4:20-cv-03664-YGR-SVK

EXPERT REPORT OF KONSTANTINOS PSOUNIS, PH.D.

June 7, 2022

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY OF OPINIONS	1
II. INTRODUCTION	12
A. Personal Background And Qualifications	12
B. Summary Of Plaintiffs' Allegations	15
C. Assignment	17
D. Facts And Data Considered	19
III. REBUTTAL TO MR. HOCHMAN	20
A. Opinion 1: Mr. Hochman's Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect	20
1. The Data At Issue Is Not Associated With A Google Account	20
2. The Data At Issue Is Stored In An Orphaned And Unidentified State	25
3. Privacy-Preserving Technical Barriers And Policies Prevent Google From Re-Identifying Logs Data	32
B. Opinion 2: Mr. Hochman's Opinions On Interception, Notice, And Deletion Of Private Browsing Information (# 4, 5, 6, 26, 31) Are Contrary To Industry Guidelines On Private Browsing	34
C. Opinion 3: Mr. Hochman's Opinions On "Private Browsing Profiles," Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate	38
1. Orphaned And Unidentified Interest Segments Are Not Cradle-To-Grave Profiles	38
2. Mr. Hochman's Claims Regarding Google's Retention Policies Show That These Policies Prevent The Creation Of Such Profiles	42
D. Opinion 4: Mr. Hochman's Assertion That Google Used Private Browsing Information To Measure Conversions (# 14) Is Misleading	43
E. Opinion 5: Mr. Hochman's Description Of Entropy (# 18) Is Incorrect	45
F. Opinion 6: Mr. Hochman's Assertions On Fingerprinting (# 9, 18) Are Misleading And Unfounded	47
1. Google Does Not Engage In Fingerprinting	48
2. Google's Internal Policies Expressly Prohibit Fingerprinting	50
3. Google's System Architecture Supports Google's Anti-Fingerprinting Policy	52
G. Opinion 7: Mr. Hochman's Proposal To Identify Class I (Chrome Class) (# 22) Is Unreasonable and Unreliable	54
1. Hochman's IP + UA Fingerprinting Method Will Not Work	56
a. IP Addresses Are Neither Unique Nor Static	56
b. User Agent Strings Are Neither Unique Nor Static	61
c. The Combination Of An IP Address And A User Agent String Is Neither	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Unique Nor Static	63
2. Pseudonymous IDs Can Not Be Used To Reliably Identify Individuals	65
a. UMA ID	65
b. PPID-Mapped Biscotti ID & Analytics User-ID	69
c. Biscotti ID	71
H. Opinion 8: Mr. Hochman’s Opinion That The “maybe_chrome_incognito” Bit Reliably Detects Incognito Traffic (# 23) Is Incorrect	72
I. Opinion 9: Mr. Hochman’s Proposal To Identify Class II (# 22) Is Unreasonable And Unreliable	75
1. Mr. Hochman’s Proposed Email Notification To All Google Account Holders Is Overly Broad And He Does Not Propose A Workable Methodology For Limiting The Notification To Class Members	77
2. Mr. Hochman’s Proposed Methodology For Limiting Class II To Private Browsing Mode Users After Notification Is Unreliable	80
J. Opinion 10: Mr. Hochman’s Proposed Methods For Identifying Class Members (# 22) Do Not–And Cannot–Account For Shared Devices Or Accounts	84
1. Mr. Hochman’s Proposed Methods Do Not Account For Shared Devices	84
2. Mr. Hochman’s Proposed Methods Do Not Account For Shared Accounts	94
IV. REBUTTAL TO MR. SCHNEIER	97
A. Opinion 11: Mr. Schneier’s Assertion That “Browsing Information Is Unique For Each User” Is Unsupported And Misleading	97
B. Opinion 12: Mr. Schneier’s Claim That “Personal Data Is Difficult To Anonymize And Easy To De-Anonymize” Is Unsupported And Is Incorrect For The Data At Issue	100
C. Opinion 13: Mr. Schneier’s Assertion That Google Has Not Taken Steps To Ensure That A User’s Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User’s Signed-Out Activity With Any Signed-In Data Is Incorrect	106
1. The Documents Mr. Schneier Cites Do Not Support His Conclusion	107
2. Additional Documents And Testimony Contradict Mr. Schneier’s Claim That Google Has Not Undertaken Steps To Prevent Joining of Signed-Out And Signed-In Data	110
3. Google’s Policy Restrictions And Pseudonymization Procedures Closely Align With Best Practices For Research Involving User Data	113
4. Google’s Security Practices Closely Align With Best Practices In The Network Security Industry	116

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

V. APPENDICES	121
A. Curriculum Vitae	121
B. Table Of Opinions	142
C. Hochman To Psounis Mapping	146
D. Psounis To Hochman And Schneier Mapping	155
E. Entropy, Entropy Bits and Fingerprinting: Formal Exposition	157
F. IP Address + User Agent Data Analysis	176
G. “Profile” Data	188
H. Sources Considered	192

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

I. EXECUTIVE SUMMARY OF OPINIONS

1. **Opinion 1: Mr. Hochman’s Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect.** In my opinion, Google cannot readily use the site activity data from users who, like putative class members, are (i) in private browsing mode, (ii) not logged into a Google Account, (iii) visiting a non-Google website that uses Google web services (the “Data At Issue”), to identify users or their devices. *See infra* [§ III.A.](#) My opinion rests on the following grounds.

- ❖ **First**, the Data At Issue is not associated with a Google Account. Google receives the data in an unidentified state, without any Google Account information. *See infra* [§ III.A.1.](#)
- ❖ **Second**, the Data At Issue is stored in an orphaned and unidentified state. The Data At Issue is either (i) keyed by a pseudonymous identifier unique to each private browsing session (*e.g.* Biscotti, Zwieback) that is not linked to an identified user or device during the session, and is not linked to anything after the session is closed, (ii) keyed by a pseudonymous identifier unique to the website publisher (PPID-mapped Biscotti, Analytics UserID) that is not linked to activity on other websites and cannot be used to identify a user, or (iii) not keyed to any identifier at all. *See infra* [§ III.A.2.](#)
- ❖ **Third**, Google’s privacy-preserving technical barriers and policies prevent the identification of users from signed-out private browsing data. *See infra* [§ III.A.3.](#)

2. **Opinion 2: Mr. Hochman’s Opinions On Interception, Notice, And Deletion Of Private Browsing Information (# 4, 5, 6, 26, 31) Are Contrary To Industry Guidelines On Private Browsing.** Mr. Hochman contends that Google could have (i) redesigned Chrome or Google’s “tracking beacons” to “limit Google’s collection of private browsing information” (Hochman # 4); (ii) notified users, at the time of collection, that it was collecting private

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

browsing information (Hochman # 5); and could have notified websites when it received private browsing information (Hochman # 6). Mr. Hochman asserts that Google “uniformly attempted to intercept all private browsing communications with non-Google websites that have a Google tracking beacon” (Hochman # 26). Finally, Mr. Hochman claims that “Google could delete from its systems records of Chrome Incognito browsing communications” (Hochman # 31). All five opinions are misguided because Mr. Hochman failed to consider industry guidelines on private browsing, including by one of the key standard-setting organizations, the W3C Technical Architecture Group (“TAG”) Observations on Private Browsing Mode, which set forth in 2019 that: (i) “the Web should be accessible in private and normal browsing modes,” and (ii) “the use of private browsing mode should not be detectable by websites.” Mr. Hochman fails to recognize that Google’s practice is consistent with these design principles. Providing a notice to websites at the time of collection or redesigning Chrome or Google scripts (Mr. Hochman’s “tracking beacons”) to “limit Google’s collection of private browsing information” would violate these fundamental principles. *See infra* [§ III.B](#).

3. Opinion 3: Mr. Hochman’s Opinions On “Private Browsing Profiles,” Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate. Mr. Hochman opines that: (1) “Google, throughout the class period, created detailed profiles ... based on the private browsing information it collected” (Hochman # 10); (2) private browsing information could be linked to information tied to the user’s Google account (Hochman # 18) and/or the user’s account with non-Google websites (Hochman # 19) but for Google’s deletion of the data (Hochman # 20). All of these opinions are inaccurate, for at least the following reasons. *See infra* [§ III.C](#).

❖ **First**, what Mr. Hochman calls “private browsing profiles” are interest segments and

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

verticals based on orphaned and unidentified data limited to single private browsing sessions, that are not joined across sessions, not joined with Google Account profiles, and not the “cradle-to-grave” profiles Plaintiffs have alleged. The activity in these so-called “profiles” is limited to discrete private browsing session activity that does not identify the user and is not linked to the user or her device after the session is closed. *See infra* [§ III.C.1.](#)

❖ **Second**, Google’s retention, deletion, and anonymization of certain data from private browsing mode sessions align with industry best practices for data security and further prevent the creation of “cradle-to-grave” profiles. Mr. Hochman’s assertion that Google should retain private browsing mode data indefinitely in order to “[aid] the process of linking (or joining) a user’s private browsing information to that user’s account” would conflict with these best practices and increase the risk of fingerprinting by bad actors. *See infra* [§ III.C.2.](#)

❖ **Third**, although Mr. Hochman cites to DBL [REDACTED] data associated with Biscotti IDs extracted from IDE cookies or mapped from PPIDs to claim that Google maintains detailed user profiles, the DBL [REDACTED] data itself shows that the purported profile is keyed to pseudonymous identifiers unique to each discrete private browsing session or unique to the website publisher. *See infra* [§ III.C.1.](#)

4. Accordingly, it is my opinion that, for the private browsing mode sessions at issue (where a user does not sign into a Google account), Google stores data from private browsing sessions keyed at most to pseudonymous identifiers unique to each private browsing session or website publisher—and sometimes to no identifier at all—subject to Google’s data retention policy. Such discrete data sets do not remotely resemble “cradle-to-grave” user

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

profiles Plaintiffs alleged in their Complaint.

5. **Opinion 4: Mr. Hochman’s Assertion That Google Used Private Browsing Information To Measure Conversions (# 14) Is Misleading.** Mr. Hochman contends that “[t]hrough Google’s server-side processes such as . . . [REDACTED] (linking a user’s signed-out Biscotti IDs across different sessions) . . . Google creates ‘a single-view of the user in all Display products’ for ad targeting and measurement purposes.” In my opinion, these assertions are misleading because they create the impression that Google maps the Biscotti ID from one of the signed-out private browsing sessions at issue in this case to the Biscotti ID from a different signed-out private browsing session. However, based on my review of documents and deposition testimony, there is no such mapping because [REDACTED] can only map a signed-in cookie (Google Accounts and ID Administration ID (“GAIA ID”)) to a signed-out cookie (*e.g.*, Biscotti) if they are in the same cookie jar. That never happens for putative class members because private browsing mode creates a new (empty) cookie jar, and Plaintiffs allege that the Class members never signed in to their Google Accounts while browsing in private mode. Accordingly, Class members could never have a signed-in and signed-out cookie in the same cookie jar, and therefore there could be no mapping. *See infra* [§ III.D.](#)

6. **Opinion 5: Mr. Hochman’s Description Of Entropy (# 18) Is Incorrect.** Mr. Hochman takes a categorical approach to defining PII, and describes entropy as a “metric for user identifiability . . . measured in number of bits of data needed to uniquely identify a person” to aid his conclusion that “IP address and User Agent ‘carries 29.8 bits’ of entropy, which is more than sufficient to uniquely identify individuals in the United States.” In my opinion, Mr. Hochman’s description is incorrect and shows a profound misunderstanding of information entropy. To support my opinions, I have provided a detailed technical discussion in

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

[Appendix E](#). In summary, there are three errors in Mr. Hochman’s opinion. *See infra* [§ III.E](#).

- ❖ **First**, contrary to Mr. Hochman’s description, entropy is not a “metric for user identifiability” but rather a measure of the uncertainty of the outcome of a process. Therefore, entropy does not—and cannot be used to—establish what specific information is necessary to uniquely identify any specific person, or how many bits of data are “more than sufficient to identify a person.” *See infra* [§ III.E](#).
- ❖ **Second**, whether IP address and User Agent (IP + UA) carry 29.8 bits, or 20 bits, or 40 bits has no bearing on user identification in this case because there is no reliable mapping from an IP + UA to an individual. Because an IP + UA pair: (i) may correspond to multiple devices and change dynamically, *see infra* [§ III.G.1](#), and (ii) may be shared by multiple users, *see infra* [§ III.G.1](#) and [§ III.J](#), there is no reliable mapping from an IP + UA pair to an individual. Therefore, an IP + UA pair will not reliably identify the person who is using the browser regardless of the entropy bits of the identifier. *See infra* [§ III.E](#).
- ❖ **Third**, entropy is an average-case metric that can provide insight into a system, but not individuals. *See infra* [§ III.E](#).

7. **Opinion 6: Mr. Hochman’s Assertions On Fingerprinting (# 9, 18) Are Misleading And Unfounded.** In his Opinion 18, Mr. Hochman opines “that Google throughout the class period and across the two classes systematically collected and stored detailed private browsing data that constitutes what is commonly understood to be sensitive fingerprinting data, which can be used to identify users and join data.” Mr. Hochman does not describe a method to distinguish “fingerprinting data” from “sensitive fingerprinting data” and, in my experience, there is no common understanding of what fingerprinting data is “sensitive.” In his Opinion 9,

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Mr. Hochman claims that “‘pseudonymous’ cookies such as Zwieback and Biscotti cookies for signed-out mode ... in fact, can be linked to users through fingerprinting information Google collects in its logs, such as IP address, user agents, websites visited, and more).” In my opinion, Mr. Hochman’s assertions on fingerprinting are misleading and unfounded for the following reasons:

- ❖ *First*, Mr. Hochman does not directly address whether Google is engaged in fingerprinting but leaves the reader with the impression that Google may be engaged in fingerprinting. *See infra* [§ III.F.1](#).
- ❖ *Second*, Mr. Hochman fails to mention that Google’s internal policies, including policies cited by Mr. Hochman, expressly prohibit fingerprinting. *See infra* [§ III.F.2](#).
- ❖ *Third*, Mr. Hochman neglects to discuss the technical barriers Google has constructed within its systems to prevent fingerprinting. *See infra* [§ III.F.3](#).

8. **Opinion 7: Mr. Hochman’s Proposal to Identify Class I (Chrome Class) (#22) Is Unreasonable and Unreliable.**

- ❖ *First*, Mr. Hochman proposes using the combination of IP address and User Agent (IP + UA) to identify class members. In short, he proposes to engage in the very fingerprinting analysis that he and Plaintiffs condemn. Even if engaging in fingerprinting were appropriate—particularly in a privacy case—this method cannot reliably identify class members because (i) it is rarely the case that a device has a unique, static IP address, and (ii) the User Agent string (UA), which contains information about the type of the browser (e.g. Chrome, Edge, Mozilla, Safari), the version of the browser, and the operating system over which the browser is running (e.g. Windows, macOS, iOS, Linux), is likely to be shared by many users. In fact, the top 10 most popular UAs

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

correspond to 26 percent of daily traffic. Therefore, the identical combination of IP address and UA may be shared by multiple class members, and by other individuals, some of whom are in the class and some of whom are not. *See infra* [§ III.G.1.](#)

❖ ***Second***, Mr. Hochman proposes using pseudonymous identifiers to identify class members. This method cannot reliably identify class members because (i) only a subset of users will even be assigned a pseudonymous identifier (*i.e.*, if they opt-in to Chrome metrics for UMA ID or if they visit websites that use PPID and/or User ID and the user signs-in to their account on those websites); (ii) UMA data is designed for aggregate analysis and is not joined with authenticated identifiers; and (iii) PPID and/or User ID are assigned by websites, purposely processed before they are sent to Google to ensure they do not contain personal identifying information, and will have the same value for users who share an account for a given website. *See infra* [§ III.G.2.](#)

❖ ***Third***, Mr. Hochman proposes using Biscotti IDs from signed-in private browsing activity to identify signed-out private browsing activity by the same user. Signed-in private browsing mode sessions cannot be joined with signed-out private browsing mode sessions via Mr. Hochman’s proposed method because (i) the Biscotti ID for signed-in activity will be different than the Biscotti ID for signed-out activity; and (ii) Mr. Hochman does not propose a way to otherwise link these two sessions together. *See infra* [§ III.G.2.c.](#)

9. **Opinion 8: Mr. Hochman’s Opinion That “maybe_chrome_incognito” Bit Reliably Identifies Incognito Traffic (# 23) Is Incorrect.** Mr. Hochman proposes using the “maybe_chrome_incognito” bit to identify web traffic by proposed class members. This method cannot reliably identify Incognito web traffic because the “maybe_chrome_incognito” bit relies

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

on the absence of the X-Client-Data header. But the X-Client-Data header may be absent for any number of reasons that have nothing to do with whether the browser was in Incognito mode. *See infra* [§ III.H](#).

10. **Opinion 9: Mr. Hochman’s Proposal To Identify Class II (# 22) Is Unreasonable And Unreliable.** Mr. Hochman proposes identifying members of Class II (Google account holders who used private browsing modes in Safari or Internet Explorer) by notifying all Google account holders based on the flawed assumption that “most” Google account holders have used a private browsing mode. Then, Mr. Hochman intends to rely on (i) self-reporting, (ii) disclosure or identification through Google records of IP address plus User Agent, and/or (iii) using Biscotti IDs from signed-in private browsing activity to identify signed-out private browsing activity by the same user. None of these methods is reliable, for at least three reasons. *See infra* [§ III.I](#).

- ❖ **First**, because private browsing modes do not save browsing history to a user’s browser, self-reporting would rely solely on the user’s memory of what websites she visited in private browsing mode on one of the browsers at issue, which is obviously unreliable. *See infra* [§ III.I.2](#).
- ❖ **Second**, for the reasons stated in my Opinion 7, neither (i) IP address plus User Agent nor (ii) attempted mapping of Biscotti IDs from signed-in private browsing mode sessions to Biscotti IDs from signed-out private browsing mode sessions method can reliably identify class members. *See infra* [§ III.G](#).
- ❖ **Third**, further highlighting the unreliability of Mr. Hochman’s method, named Plaintiffs’ data shows that each potential class member’s Google account may be associated with multiple combinations of IP address plus User Agent. *See infra* [§ III.I.2](#).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

11. **Opinion 10: Mr. Hochman’s Proposed Methods For Identifying Class Members (# 22) Do Not—And Cannot—Account For Shared Devices.** Even if Mr. Hochman’s proposed methods for class member identification could reliably link the private browsing activity at issue to a specific device or account identifier (and as I explain in my Opinions 7 through 9, it cannot), the method would still not be able to identify class members because it does not address the issue of device and Google account sharing.

❖ **First**, there is a significant body of research (including research by one of Plaintiffs’ other experts, Mr. Schneier) showing that sharing of devices is common, and likely even more common among private browsing users because Incognito mode and other private browsing modes’ primary use case is to provide on-device privacy for shared devices. *See infra* [§ III.J.1.](#)

❖ **Second**, users may share Google accounts and accounts for non-Google websites, such as for websites that require a paid subscription. *See infra* [§ III.J.2.](#)

❖ **Third**, in conjunction with Mr. Hochman’s proposal to send email notifications based on device or account, his failure to address shared devices and accounts may put those private browsing users who seek to keep that activity private from users who share those devices and/or accounts at risk (*e.g.*, an abused spouse or child doing research on how to find help). *See infra* [§ III.J.](#)

12. **Opinion 11: Mr. Schneier’s Assertion That “Browsing Information Is Unique For Each User” Is Unsupported and Misleading.** Mr. Schneier purports to support his opinion with studies and documents that do not actually support his opinion because they do not conclude that browsing information is in fact unique for each user, as Mr. Schneier claims. His opinions are particularly misleading for the Data At Issue because the studies and

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

documents Mr. Schneier cites rely on data and methodologies that are inapplicable to the data in this case. *See infra* [§ IV.A.](#)

13. **Opinion 12: Mr. Schneier’s Claim That “Personal Data Is Difficult to Anonymize and Easy to De-anonymize” Is Unsupported And Is Incorrect For The Data At Issue.** As explained further below, Mr. Schneier’s assertion is unsupported by the sources he cites and incorrect for the Data At Issue for the following reasons:

- ❖ *First*, the examples Mr. Schneier cites as examples of purported difficulties associated with anonymization analyze data that is readily distinguishable from the Data at Issue—for example, surveillance camera footage of Israeli assassins, birth dates, ZIP codes, and voter registration databases. *See infra* [§ IV.B.](#)
- ❖ *Second*, the Google documents that Mr. Schneier characterizes as “admissions” do not support his conclusion that “Google can connect individuals to private browsing sessions.” *See infra* [§ IV.B.](#)
- ❖ *Third*, Mr. Schneier’s assertion that “data is difficult to anonymize and easy to de-anonymize” is incorrect for the Data At Issue because my review of Google documents and testimony produced in this case confirms that Google has been successful in its efforts to anonymize the Data At Issue and prevent its re-identification. *See infra* [§ IV.B.](#)

14. **Opinion 13: Mr. Schneier’s Assertion That Google Has Not Taken Steps To Ensure That A User’s Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User’s Signed-Out Activity With Any Signed-In Data Is Incorrect.** In short, Mr. Schneier’s opinion is incorrect for the following reasons:

- ❖ *First*, the documents that Mr. Schneier cites do not support his conclusion that Google

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

“has not taken steps to ensure that a user’s choice to sign out of a Google account will prevent Google from associating the user’s signed-out activity with any signed-in data;” in fact, they support the opposite conclusion because they outline the extensive steps Google takes to maintain strict separation between signed-in and signed-out data. *See infra* [§ IV.C.1.](#)

❖ ***Second***, additional documents and testimony in this case further demonstrate the steps that Google has taken to ensure that a user’s choice to sign out of a Google account will prevent Google from associating the user’s signed-out activity with any signed-in data. *See infra* [§ IV.C.2.](#)

❖ ***Finally***, upon review of Google’s security and pseudonymization policies and practices, they closely align with best practices for research involving user data and with best practices in the network security industry. *See infra* [§ IV.C.3.](#)

15. For an overview of my opinions, please refer to [Appendix B. Table Of Opinions](#). For tables mapping Mr. Hochman’s opinions to my report and my report to his opinions, please see [Appendix C. Hochman to Psounis Mapping](#) and [Appendix D. Psounis To Hochman Mapping](#).

II. INTRODUCTION

A. Personal Background And Qualifications

16. I am a Professor and Associate Chair of Electrical and Computer Engineering and Professor of Computer Science at the University of Southern California. I joined the University of Southern California in 2003, after completing my PhD at Stanford University as a Stanford Graduate Fellow.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

17. My professional career has spanned more than 20 years. As set forth in [Appendix A. Curriculum Vitae](#), I have extensive experience in the field of networked distributed systems, including the Internet and the world wide web, content-delivery networks, data centers and cloud computing, and wireless mobile networking systems.

18. I have published more than 100 technical papers in these fields, which have been cited tens of thousands of times. I have also been awarded numerous grants and significant funding from the government and industry leaders to advance these fields. As a result, I have been named an Institute of Electrical and Electronics Engineers (IEEE) Fellow, the highest grade of membership, and a Distinguished Member of the Association of Computing Machinery (ACM) for my contributions to the theory and practice of networked, distributed systems.

19. Throughout my career, I have analyzed, designed, and developed efficient, privacy-preserving networked distributed systems for the Internet and the Web, content-delivery networks, data centers and cloud systems, and wireless mobile networking systems. As such, I have acquired extensive expertise in the analysis and development of those systems and associated products.

20. I have extensive experience with and made contributions specifically towards designing efficient, privacy-preserving networked distributed systems, as established by my funding and publication record. Over the last decade, I have received multiple awards from the National Science Foundation (NSF), the leading governmental agency for funding computer engineering and computer science research, to work on privacy-preserving architectures for networked distributed systems, including the world wide web, mobile systems, and spectrum sharing systems. For example, I am a lead Principal Investigator (PI) of the NSF Secure and Trustworthy Cyberspace Frontiers (SaTC Frontiers) grant on Protecting Personal Data Flow on

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

the Internet and the sole lead PI of the NSF grant on Spectrum Sharing Systems for Wireless Networks: Performance and Privacy Challenges. During my professional career of more than 20 years I have also received multiple awards from NSF and industry leaders, to work on network routing, for example, I was the sole lead PI on the NSF grant on Efficient Routing in Delay Tolerant Networking.

21. I have published several papers in selective academic journals and conferences on protecting the privacy of user and system data. Recent work of mine in the area of privacy-preserving distributed systems concerns the protection of the privacy of personal data during web browsing and during usage of popular applications and systems, touching upon implications of cookies and fingerprinting.¹ In addition, I have a large body of publications in the most selective academic journals and conferences related to network routing and IP addresses.² I have also been the faculty in charge of the entire networking curriculum at the Electrical and Computer Engineering department at USC for more than a decade and teach networking classes as well as probability theory classes which cover entropy and other related concepts to graduate students yearly.

22. I also have extensive practical experience with networked distributed systems designed and developed to operate in the World Wide Web deployed over the Internet. For example, I was the Technology Architect for Fineground Networks (later acquired by Cisco

¹ See, e.g., M. Clark and K. Psounis, "Optimizing Primary User Privacy in Spectrum Sharing Systems," IEEE/ACM Transactions on Networking, <https://ieeexplore.ieee.org/document/8985324> (Apr. 2020); J. Zhang, K. Psounis, M. Zaroon, and Z. Shafiq, "HARPO: Learning to Subvert Online Behavioral Advertising," NDSS, <https://web.cs.ucdavis.edu/~zubair/files/harpo-ndss2022.pdf> (Apr. 2022).

² See, e.g., T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-copy Case," IEEE/ACM Transactions on Networking, <https://cpb-us-e1.wpmucdn.com/sites.usc.edu/dist/b/364/files/2019/05/multiton.pdf> (Feb. 2008); T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Single-Copy Case," IEEE/ACM Transactions on Networking, <https://ee.usc.edu/netpd/assets/001/51984.pdf> (Feb. 2008).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Systems), where I designed and developed systems to accelerate the delivery of content over the World Wide Web via web proxy caches and content delivery networks. Also, I was a co-founder of SpaceMUX (whose IP was later acquired by Quantenna Communications), where I designed and developed systems to increase the speed of wireless networked systems. In addition, for over 20 years both at Stanford University and at the University of Southern California I have designed and implemented efficient algorithms, protocols and systems for web caching, web farms, and web browsing, and I have consulted for industrial leaders, and produced prototype systems. I have also applied for and been granted numerous patents, e.g. “Method and System for Class-based Management of Dynamic Content in a Networked Environment”³ about dynamic web content, owned by Cisco Systems.

23. In sum, I have extensive experience in and familiarity with the fields of networked distributed systems including the Internet and the world wide web, content-delivery networks, data centers and cloud computing, and wireless mobile networking systems, and extensive experience and contributions towards the analysis, design and implementation of efficient routing and privacy-preserving architecture for such systems.

24. My C.V. is attached as [Appendix A. Curriculum Vitae](#).

25. I am being compensated at the rate of \$600 per hour for my work on this case. My compensation is not contingent upon my findings, the testimony I may give, or the outcome of this litigation.

B. Summary Of Plaintiffs’ Allegations

26. I understand that Plaintiffs, Chasom Brown, William Byatt, Jeremy Davis, Christopher Castillo, and Monique Trujillo, allege that Google violated various statutes and

³ [U.S. Patent No. 7,603,483B2](#) (issued Oct. 13, 2009).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

privacy laws since June 1, 2016 (the “Class Period”).⁴ Specifically, Plaintiffs allege that Google improperly received or collected data from browsers when (i) a Google Account holder not signed into their Google Account, (ii) uses a browser in private browsing mode, (iii) to visit a non-Google website containing “Google tracking or advertising code.”⁵ The Data At Issue includes the following:

- “The ‘GET request’ sent from the user’s computer to the website;”
- “The IP address of the user’s connection to the internet;”
- “Information identifying the browser software that the user is using, including any ‘fingerprinting’ data;”
- “Any ‘User-ID’ issued by the website to the user, if available;”
- “Geolocation of the user, if available;” and
- “Information contained in ‘Google cookies,’ which were saved by the user’s web browser on the user’s device at any prior time.”⁶

27. Plaintiffs allege that Google uses “fingerprinting”⁷ techniques to (i) intercept and collect the Data At Issue⁸ and (ii) build “its profile of users (including Plaintiffs and class members).”⁹ In their Third Amended Complaint, Plaintiffs claim that “Google Identifies You with ‘Fingerprinting’ Techniques.”¹⁰ Plaintiffs further allege “Google Ad Manager is set up to automatically retarget a user based on information that Google has previously collected”

⁴ Third Amended Class Action Complaint, *Chasom Brown, et al., v. Google LLC*, United States District Court Northern District of California, Febr. 3, 2022, Dkt. 395-2 (“Complaint”) ¶ 2.

⁵ *Id.* ¶ 192.

⁶ *Id.* ¶ 63.

⁷ In the context of browser communications, fingerprinting refers to the combination of various bits of information to probabilistically identify a browser. Probabilistic identification of a browser through fingerprinting is different from the deterministic identification of an individual through an account log-in. For more on fingerprinting *see infra* [§ III.F](#), [§ III.G.1](#), [Appendix E](#).

⁸ Complaint ¶ 8.

⁹ *Id.* ¶ 100.

¹⁰ *Id.* at 31.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

through “Google’s fingerprinting.”¹¹ Plaintiffs allege that Google engages in “[c]oncurrent retargeting of ads by matching private browsing user data with other user data.”¹²

28. Plaintiffs allege that Google “gained a complete, cradle-to-grave profile of users” by correlating and aggregating data associated with:

- **Analytics User-ID.** (“Google is able to associate the data collected from users in “private browsing mode” with specific and unique user profiles through Google Analytics User-ID.”)
- **Google Cookies.** (“Information collected from Google Cookies, which includes identifying information regarding the user from private browsing sessions and non-private browsing sessions, across multiple sessions[.]”)
- **Fingerprinting Data.** (“Identifying information regarding the consumer from various Google fingerprinting technologies that uniquely identify the device, such as X-Client-Data Header, GStatic, and Approved Pixels[.]”)
- **Geolocation Data.** (“Geolocation data that Google collects from concurrent Google processes and system information, such as from the Android Operating System[.]”)
- **IP Address.** (“The IP address information, which is transmitted to Google’s servers during the private and non-private browsing sessions. Google correlates and aggregates all of this information to create profiles on the consumers.”)¹³

29. Plaintiffs propose two classes:

- **Class 1** – “All Chrome browser users with a Google account who accessed a non-Google website containing Google tracking or advertising code using such a browser and who were (a) in ‘Incognito mode’ on that browser and (b) were not logged into their Google account on that browser, but whose communications, including identifying information and online browsing history, Google nevertheless intercepted, received, or collected from June 1, 2016 through the present (the ‘Class Period’).”¹⁴
- **Class 2** – “All non-Chrome browser users with a Google account who accessed a non-Google website containing Google tracking or advertising code using any such browser and who were (a) in ‘private browsing mode’ on that browser, and (b) were not logged into their Google account on that browser, but whose communications, including

¹¹ *Id.* ¶ 81.

¹² *Id.* at 25.

¹³ *Id.* ¶ 93.

¹⁴ *Id.* ¶ 192.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

identifying information and online browsing history, Google nevertheless intercepted, received, or collected from June 1, 2016 through the present (the ‘Class Period’).”¹⁵

C. Assignment

30. I have been engaged in this matter by counsel for Google LLC (“Google”) to evaluate the April 15, 2022 Expert Report of Jonathan E. Hochman (“Hochman Report”)¹⁶ and the April 15, 2022 Expert Report of Bruce Schneier (“Schneier Report”).¹⁷

31. In particular, counsel has asked me to analyze and respond to the following Opinions in the Hochman report.

- **A.** “Google Interception: Throughout the class period, Google intentionally intercepted private browsing communications between users and non-Google websites while those communications were in transit.” (Hochman ¶¶ 78-133.)
- **B.** “User Notification & Choice: Throughout the class period, Google collected this private browsing information without notifying users or offering a choice at the time of or in connection with any of the interceptions.” (Hochman ¶¶ 134-135).
- **C.** “Website Notification & Choice: Throughout the class period, Google collected this private browsing information without notifying websites at the time of or in connection with any of the interceptions or providing websites with a choice.” (Hochman ¶¶ 136-137).
- **D.** “Google Storage: Throughout the class period, Google stored private browsing information in many Google data sources, sometimes permanently.” (Hochman ¶¶ 138-165).
- **E.** “Google Use: Throughout the class period, Google exploited private browsing information to generate revenue for Google by creating profiles, serving ads, tracking conversions, and in other ways that benefited Google.” (Hochman ¶¶ 166-222).
- **F.** “Google Joinability: Throughout the class period, Google collected and stored private browsing information in ways that can be joined to other Google user information, but Google withheld and destroyed data that would be relevant to further assessing and demonstrating that joinability.” (Hochman ¶¶ 223-258).

¹⁵ *Id.* ¶ 192.

¹⁶ Apr. 15, 2022 Expert Report of Jonathan E. Hochman and Appendices and Exhibits Attached Thereto.

¹⁷ Apr. 15, 2022 Expert Report of Bruce Schneier and Appendices and Exhibits Attached Thereto.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- **H.** “Class Member Identification: Throughout the class period, Google collected private browsing information in ways that can be used to identify class members, though Google also withheld and destroyed data relevant to that identification.” (Hochman ¶¶ 262-307).
- **I.** “Attempt: Google’s attempted interception and collection uniformly impacted all class members.” (Hochman ¶¶ 308-317).

32. Counsel has also asked me to analyze and respond to the following specific statements in the Schneier Report:

- “Browsing Information Is Unique For Each User.”¹⁸
- “Personal Data Is Difficult to Anonymize and Easy to De-anonymize.”¹⁹
- “Google has not taken steps to ensure that a user’s choice to sign out of a Google account will prevent Google from associating the user’s signed-out activity with any signed-in data.”²⁰

D. Facts And Data Considered

33. I have reviewed the Hochman Report, exhibits, and appendices, and sources relied on, as well as the Schneier Report. I have also requested and reviewed Google log data provided pursuant to the Special Master process, Google documents and discovery responses produced in this case, including regarding Google’s logging policies and infrastructure, its anti-fingerprinting policies and infrastructure, all transcripts of Google deposition witnesses in this case discussing Google’s logging policies and infrastructure, issues surrounding fingerprinting and IP addresses, publicly available data and reports, and existing technical and academic research. The sources I considered in forming my opinions are identified in this report, the accompanying exhibits, and are listed in [Appendix H. Sources Considered](#).

34. Should additional relevant documents or information be made available to me, I reserve the right to supplement my opinions as appropriate.

¹⁸ Schneier at 27.

¹⁹ Schneier at 41.

²⁰ Schneier ¶ 205.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

III. REBUTTAL TO MR. HOCHMAN**A. Opinion 1: Mr. Hochman’s Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect**

35. I disagree with Mr. Hochman’s “opinion that Google can readily use the data at issue in this lawsuit (collected from private browsing during the class period) to identify users and their devices.”²¹ In my opinion, Google cannot readily use the collected site activity data from (i) browsers in private browsing mode, (ii) with no user logged in to a Google Account, (iii) visiting a non-Google website that uses Google web services (the “Data At Issue”), to identify users or their devices for the following reasons:

- ❖ **First**, the Data At Issue is not associated with a Google Account. Google receives the data in an unidentified state, without any Google Account information. See *infra* [§ III.A.1.](#)
- ❖ **Second**, the Data At Issue is stored in an orphaned and unidentified state. The Data At Issue is either (i) keyed by a pseudonymous identifier unique to each private browsing session (Biscotti, Zwieback), (ii) keyed by a pseudonymous identifier unique to the website publisher (PPID-mapped Biscotti, Analytics UserID), or (iii) not keyed to any identifier. See *infra* [§ III.A.2.](#)
- ❖ **Third**, Google’s privacy-preserving technical barriers and policies prevent the identification of users from signed-out private browsing data. See *infra* [§ III.A.3.](#)

1. The Data At Issue Is Not Associated With A Google Account

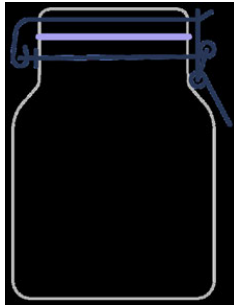
36. Mr. Hochman recognizes that, according to Plaintiffs’ Class definition, Class members are not signed-in to Google Accounts when Google receives the data. He states that “for the private browsing information at issue in this case . . . class members are not signed into

²¹ Hochman ¶ 227.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

any Google account”²² and “[w]hen a user starts a new private browsing session, they are automatically signed-out of any signed-in Google account.”²³ In addition, he states that “Google distinguishes between signed-in and signed-out modes by using different cookies.”²⁴ Finally, Mr. Hochman states that “Google stores . . . data in different logs depending on the signed-in vs. signed-out mode.”²⁵ Yet, despite these acknowledgments, he goes on to conclude that “Google can readily use the data at issue in this lawsuit (collected from private browsing during the class period) to identify users and their devices.”²⁶

37. Based on my experience researching browser communications, review of documents upon which Mr. Hochman relied, review of deposition testimony and data produced in this litigation, I conclude that Mr. Hochman’s opinion is incorrect. That the Data At Issue is not associated with a Google Account is easily illustrated:

<p>38. This empty jar represents a browser’s local cookie storage on a brand new browser.</p>	
---	---

²² Hochman ¶ 156.

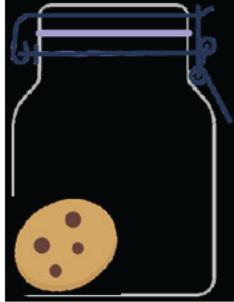


²³ Hochman ¶ 161.

²⁴ Hochman ¶ 160 (“Google uses GAIA cookies for signed-in mode and what Google calls ‘pseudonymous’ cookies such as Zwieback and Biscotti cookies for signed-out mode.”).


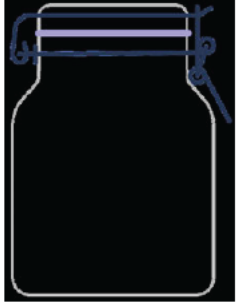
²⁵ *Id.* (“[I]n signed-out mode, Google logs data to Biscotti logs (B logs) for display ads data and Zwieback logs for search ads data.”).

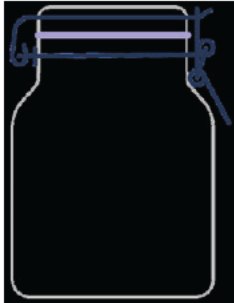
²⁶ Hochman ¶ 227.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<p>39. If a user signs in to a Google Account on that browser (e.g., by logging in at https://mail.google.com), a cookie associated with that Google Account is set on the browser. This cookie contains an encrypted ID, called a GAIA, which is used to identify the specific Google Account. This cookie remains on the browser while the user is signed in to their Google Account.</p>	 <ul style="list-style-type: none"> GAIA ID for USER@gmail.com
<p>40. Over time, depending on the browser's cookie settings and the web pages visited, other cookies may be set. For example, if the browser is used to visit a non-Google web page on which the web page publisher has installed Google Ad Manager scripts, a cookie containing a unique encrypted Biscotti ID may be set.</p>	 <ul style="list-style-type: none"> GAIA ID for USER@gmail.com Biscotti ID # 1 (2515782358150873158)
<p>41. If the browser is used to visit a web page Google owns and operates, a cookie containing a unique encrypted Zwieback ID may be set. All of these cookies can exist on a browser at the same time.</p>	 <ul style="list-style-type: none"> GAIA ID for USER@gmail.com Biscotti ID # 1 (2515782358150873158) Zwieback ID # 1 (0xf5c156f144b9aa7f)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<p>42. When a private browsing mode is enabled on a browser (<i>e.g.</i>, Incognito on Chrome), the browser creates a new cookie jar for the duration of that private browsing session. That cookie jar does not contain any of the previously set cookies, or encrypted IDs stored in those cookies.²⁷</p>	 <ul style="list-style-type: none"> • GAIA ID for USER@gmail.com • Biscotti ID # 1 (2515782358150873158) • Zwieback ID # 1 (0xf5c156f144b9aa7f) 
--	--

<p>43. Because class members (Class I and Class II) are not signed into any Google account for the private browsing sessions at issue in this case,²⁸ Google never receives the Data At Issue with a GAIA ID, and the data is not associated with a Google Account.</p>	
--	--

44. Mr. Hochman states that “Google stores ... data in different logs depending on the signed-in vs. signed-out mode.”²⁹ These “signed-out mode” logs store data keyed to an unauthenticated identifier—not a GAIA.³⁰ Data stored in these logs is therefore not keyed to a specific user’s identity.³¹ This data is unauthenticated (*i.e.*, not associated with a specific user’s

²⁷ Hochman ¶ 161. (“When a user starts a new private browsing session, they are automatically signed-out of any signed-in Google account.”).

²⁸ Hochman ¶ 156.

²⁹ Hochman ¶ 160.

³⁰ See GOOG-CABR-00059431, at -431 [REDACTED]

³¹ See GOOG-CABR-00073880, at -881 (“Unauthenticated User Data is User Data that is not tied to a signed-in user.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

identity), and Google implements policies and technical controls to prevent its re-identification.³²

45. In 2016, Google undertook a worldwide effort to “give signed-in users better control and transparency for their ads experience.”³³ The █████ 2.0 program introduced a schism in how Google’s systems identified users. Prior to █████ 2.0, Google’s ad systems used an unauthenticated Biscotti ID, or doubleclick ID, as the sole identifier, whether or not users were signed in to their Google Account. With the introduction of █████ 2.0, Google’s ad systems █████
 █████³⁴ █████ 2.0 introduced similar changes to Google Analytics to █████
 █████³⁵ With the introduction of GAIA-keyed storage and processing, Google implemented changes to its logging pipelines to maintain and enforce separation between GAIA-keyed authenticated logs and Biscotti-keyed unauthenticated logs.³⁶

³² See *id.* at -882 █████

”).

³³ GOOG-CABR-04750983, at -984.

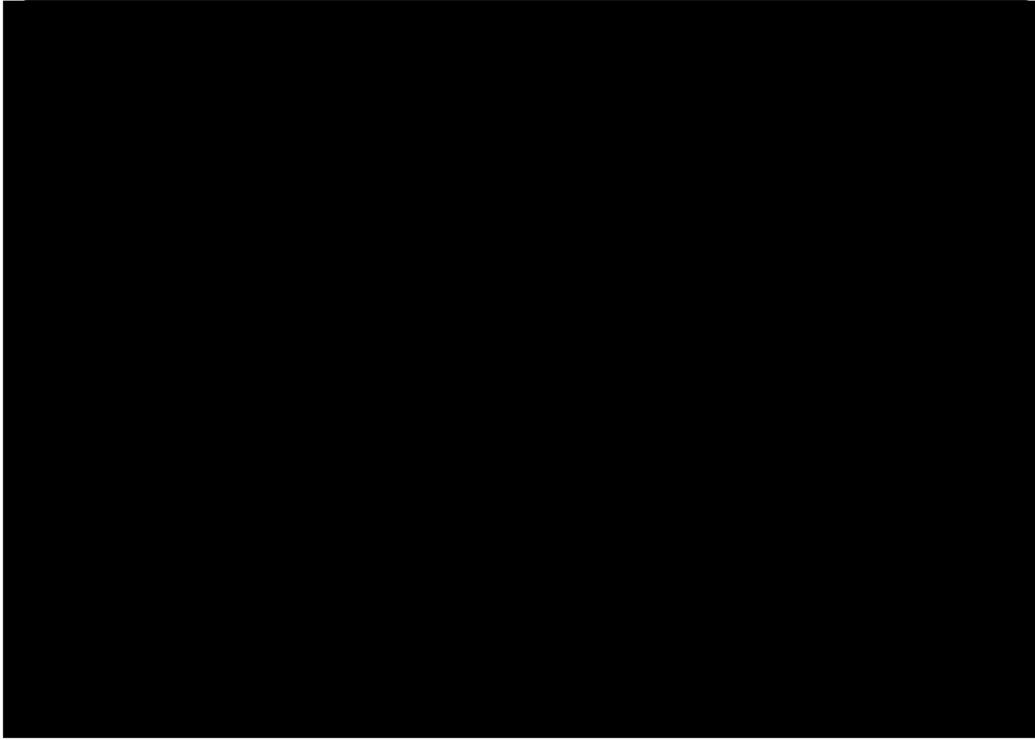
³⁴ GOOG-CABR-00101695, at -696.

³⁵ *Id.*

³⁶ *Id.* at -697 █████

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

46. The serving stack for Display Ads uses encryption to maintain the separation between Biscotti and GAIA-keyed data and limit the risk of joinability. *See* GOOG-CABR-04717111, at -138:

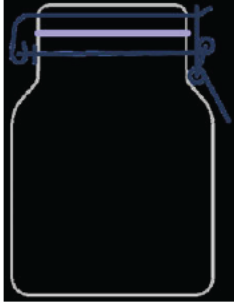
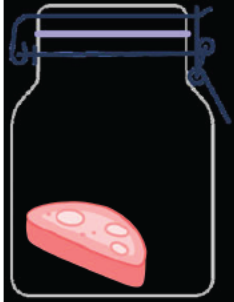



47. In my opinion, because (i) the Data At Issue is not received with a GAIA ID and (ii) Google maintains separation between GAIA-keyed data and non-GAIA-keyed data, this Data is not associated with a Google Account.

2. The Data At Issue Is Stored In An Orphaned And Unidentified State


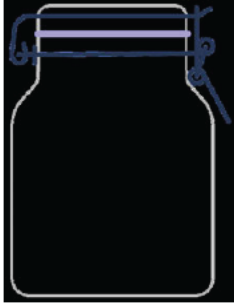


48. While Google does not receive the Data At Issue with a GAIA ID, it can receive the Data At Issue with the pseudonymous Biscotti and Zwieback IDs discussed above. That those pseudonymous IDs are unique to each private browsing session can easily be illustrated with our prior example.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<p>49. When a user enables private browsing mode on a browser (<i>e.g.</i> starts an Incognito session on Chrome), a new cookie jar is created that is only in operation until the end of the private browsing session.</p>	
<p>50. When a user visits a non-Google web page on which the web page publisher has installed Google Ad Manager scripts, and the browser permits third-party cookies to be set, a cookie containing a unique encrypted Biscotti ID may be set. That Biscotti ID is unique to the specific private browsing session. It has no relation to, and is not joined with, any cookies set during regular (what Hochman calls “non-private” browsing) mode, or even a Biscotti ID set during a different private browsing session.³⁷</p>	 <ul style="list-style-type: none"> • Biscotti ID # 2 (2483711983363751938)
<p>51. When a user visits a web page Google owns and operates, a cookie containing a unique encrypted Zwieback ID may be set. The Zwieback ID is unique to the private browsing session. It has no relation to, and is not joined with, any cookies set during regular mode, or even a Zwieback ID set during a different private browsing session.</p>	 <ul style="list-style-type: none"> • Biscotti ID # 2 (2483711983363751938) • Zwieback ID # 2 (0x43b763007d1ff50d)

³⁷ See Berntson June 16, 2021 Tr. 279:1-11 (“When you create an incognito session, the incognito session starts off with an empty cookie jar that is not shared with anything else on your device. So yes, when an ad request is made from an incognito session for the first time, there is no Biscotti, so it will generate a brand new Biscotti that has no mapping to anything else.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<p>52. When a user closes the private browsing session, the cookies that were set during that private browsing session are deleted from the browser.</p>	
<p>53. When a user opens a new private browsing session, a new cookie jar is created that is only in operation until the end of the private browsing session.</p>	
<p>54. If a user returns to the same web pages they visited in the prior private browsing sessions, new cookies containing new Zwieback and Biscotti IDs will be set.</p>	 <ul style="list-style-type: none"> • Biscotti ID # 3 (2464002382754152675) • Zwieback ID # 3 (0x6e21bea56616853b)
<p>55. When a user closes the private browsing session, the cookies that were set during that private browsing session are deleted from the browser.</p>	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

56. As a consequence of the interaction shown above, when Google receives the Data At Issue with a Biscotti ID or a Zwieback ID, that ID is unique to the specific private browsing session and the Data At Issue is stored in an orphaned state.

57. As former Google employee Rory McClelland explained at his deposition:

For example, [if] a user who launched an Incognito window . . . navigated to Google.com and undertook a search, the Google.com web service in this instance would create a new cookie for that user, that user presenting as a whole new user that has never been seen before due to the nature of Incognito mode. They are given a new cookie that holds the unique identifier for that profile. That cookie is held in memory on the user's computer because it's in Incognito mode. For the duration of that search, session data is logged against the ID that represents that profile on the server. When the user closes the last Incognito tab or window, locally, that cookie is erased from memory . . . That profile, with its corresponding data will become—best way of explaining, orphaned, so it will continue to exist, but it would never be supplemented with any additional data.

McClelland Feb. 18, 2022 Tr. 80:12-81:11.³⁸

58. Based on my experience in the industry and review of data and documents for this case, I agree with Mr. McClelland's description. When a user does not sign into a Google account in Incognito mode (or private browsing mode on Safari, Internet Explorer, or Edge), any cookies (and the identifiers contained therein) that are set will be unauthenticated (*i.e.*, not associated with a specific user's identity by Google). When the private browsing session is terminated, the cookies are deleted from the browser.

59. Google can also receive the Data At Issue with a pseudonymous first-party ID called PPID.³⁹ Notably, Mr. Hochman has proposed in published research that the use of "local patient identifiers (LPIDs) that can be used to identify the medical records of a given patient within the context of a single health care provider" provides an effective way to facilitate

³⁸ See also Berntson June 16, 2021 Tr. 279:1-11.

³⁹ Google Ad Manager Help, "About publisher provided identifiers," <https://perma.cc/H8L4-FYTM>, (last visited June 6, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

sharing of medical records for research purposes while protecting patients' privacy.⁴⁰ This LPID is "obtained from a patient's PII via a one-way cryptographic function . . . [that] prevents the local identifier from being reverse engineered to obtain PII."⁴¹ PPIDs function in the same way as Mr. Hochman's proposed LPIDs. The PPID is a "[p]ublisher provided identifier [that] allows publishers to send Google Ad Manager an identifier for use in frequency capping, audience segmentation and targeting, sequential ad rotation and other audience-based ad delivery controls across devices."⁴² The PPID is set by publishers and Google requires publishers to send the PPID in a form to Google that does not contain any identifying information: "This identifier [PPID] . . . must be hashed, anonymous, and must not contain any personal information, third-party identifiers or device IDs."⁴³

60. Google does not link PPIDs to other identifiers and does not use PPIDs for Google's own purposes.⁴⁴ They are controlled by the publisher and are specific to individual publishers and their websites.⁴⁵ Because publishers are in full control, not all publishers use

⁴⁰ M. Fischer, J. Hochman, and D. Boffa, "Privacy-Preserving Data Sharing for Medical Research," International Symposium on Stabilization, Safety, and Security of Distributed Systems, <https://cpsc.yale.edu/sites/default/files/files/TR1558.pdf> (November 17–20, 2021), at 5.

⁴¹ *Id.*

⁴² GOOG-CABR-00021811, at -811.

⁴³ *Id.*

⁴⁴ Berntson Mar. 18, 2022 Tr. 119:17-23 ("In addition, from an implementation perspective, for the PPID, it's never ever linked to any other ID. [Its] representation changes externally and then internally, but it's all only ever one ID that is never joined to any other ID in any business case."); *id.* at 123:24-124:15 ("PPID doesn't help with conversions because the ID only exists on the publisher's site. And once a user has, say, clicked on an ad and gone to an advertiser site, the PPID is not relevant anymore. And in fact, I'll make a linkage here. One of the uses of modeled conversions is specifically when we don't have an ID that spans cross-site. And in those cases, PPID is a good example. We can't measure conversions using a PPID because it only exists on the publisher's site, not on the advertiser's site. And when PPID is the ID that is used, it's -- for example, if -- if a user has disabled third-party cookies, there is no Biscotti. And in those cases, there is no cross-site identifier that can be used to infer conversions.").

⁴⁵ *Id.* at 167:3-14 ("So [a] design attribute to our products is to put the publisher in control of the data that they want used in their system. That's different than the controls we have relating to the e-privacy directive, because for that, even if the data is present, we wouldn't use it. The

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

PPIDs,⁴⁶ and different publishers may provide Google a hashed PPID value that is the same for the same user.⁴⁷

61. When Google receives a PPID value from a publisher, Google hashes it further,⁴⁸ and then uses a table to map it to a random ID.⁴⁹ The random ID is called a PPID-mapped Biscotti ID because “‘Biscotti’ was originally named for the internal system [Google] use[s] to generate random numbers.”⁵⁰ It is not the same Biscotti ID as stored in a cookie on the user’s browser.⁵¹ The mapping table records the double-hashed PPID value, and the hashing is one-way so there is no way to retrieve the original PPID input.⁵²

62. For the Data At Issue, the PPID-mapped Biscotti ID is only written to logs that are not associated with a Google Account. For cases when a user is signed into their Google Account and the publisher provides a PPID to Google, the PPID-mapped Biscotti ID is stored in an encrypted format to prevent logged-out private browsing traffic from being joined with a user’s Google Account.⁵³

second set of controls is a publisher saying, ‘I don’t want you to introduce that data’ or ‘I’m not going to provide it to you.’ So those are two different pathways where there could be limitations or restrictions placed on whether or not we have access to data.”).

⁴⁶ *Id.* at 166:16-20 (“[I]f a publisher chooses to not use PPID, they wouldn’t provide PPID, and so therefore there would be no PPID. And so, it is a publisher choice as to whether the PPID [is] available in the first place for our systems”).

⁴⁷ *Id.* at 121:20-24 (“[I]f two different publishers have the same e-mail address and they happen to use the same hashing function, it is possible that two different publishers can pass us what -- a PPID value that’s the same for the same user.”).

⁴⁸ *Id.* at 120:12-13 (“when we receive this [PPID] value, we hash it”).

⁴⁹ *Id.* at 120:22-121:1 (“And when this comes in the first time, we generate a new random number, and we associate that random number -- which is an integer -- and we store it alongside what we get from the publisher.”).

⁵⁰ *Id.* at 122:11-15.

⁵¹ *Id.* at 122:15-24 (“And that internal system that generated a random integer is also what we happen to call the cookie that we publish client-side. . . . That mapped Biscotti ID is never present external to our systems.”).

⁵² *Id.* at 121:8-11 (“[T]hey hash it into, say, another string, which is an opaque reputation with a one-way mapping from the prior string.”).

⁵³ GOOG-CABR-00058751, at -756.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

63. Similarly, Google Analytics User-ID is also a first-party ID generated, assigned, and used by Google Analytics customers.⁵⁴ A publisher who is a Google Analytics customer can provide a hash of their signed-in User ID to Google; that signed-in User ID is called an Analytics User-ID.⁵⁵ Because publishers have full control of Analytics User-ID, not all publishers choose to use Analytics User-ID (some publishers do not have a signed-in user base and thus cannot use Analytics User-ID⁵⁶), and different publishers may provide Google a hashed Analytics value that is the same for the same user.

64. In my opinion, Google can not readily use the Data At Issue to identify users because the Data is stored in an orphaned and unidentified state.

⁵⁴ Google Analytics Help, “About User-ID views,” <https://perma.cc/9MGV-G8XP> (last visited June 6, 2022).

⁵⁵ Berntson June 16, 2021 Tr. 235:16-236:2 (“If a publisher has a signed-in user base, they can provide a hash of say their signed-in User ID, which means that when they’re using Google Analytics to look at a user’s activity on a mobile device and on a desktop, et cetera, because the publisher is providing a stable identifier across those devices, they can have the view of all the activity associated with one user across devices.”).

⁵⁶ *Id.* at 240:12-13 (“Many publishers don’t have a signed-in user base”); Chung Mar. 10, 2022 Tr. 132:14-17 (“[T]he customer would have to implement the user ID feature, which is in the -- the large minority of Google Analytics’ customers and sites use user ID and send event data in with user ID.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

3. Privacy-Preserving Technical Barriers And Policies Prevent Google From Re-Identifying Logs Data

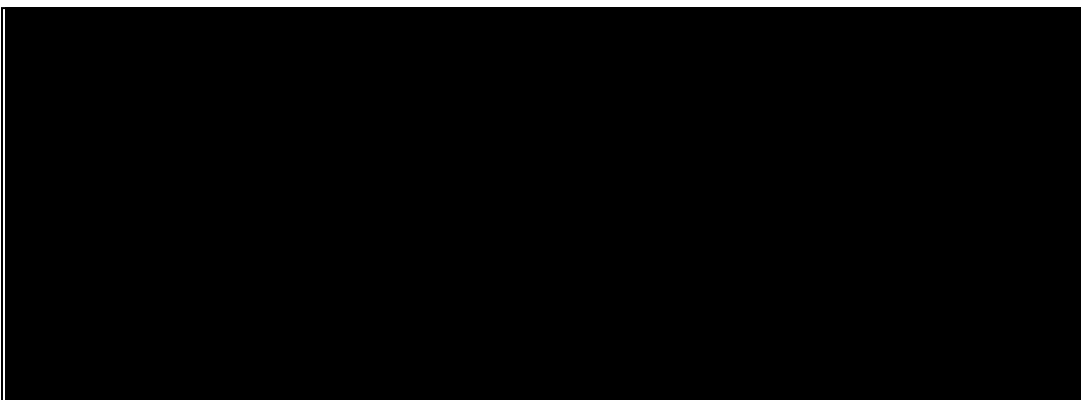
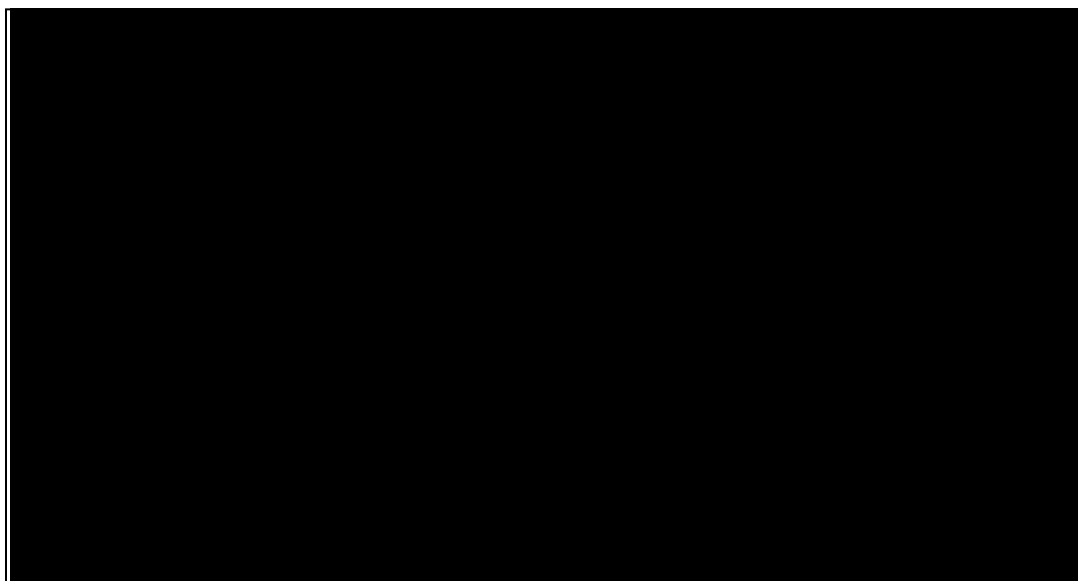
65. In my opinion, Google can not readily identify users from the Data At Issue due to technical barriers and policies designed to prevent such identification. Google has strict policies against re-identifying logs data. *See, e.g.*, GOOG-BRWN-00029004, at -006:



CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

66. Additional policies preventing re-identification include Google's policies against fingerprinting are discussed in [§ III.F.2](#). Google also performs "cookie scrubbing," which is a process where [REDACTED]

57



67. Additional relevant technical barriers to re-identification include [REDACTED] architecture and encryption (*infra* [§ III.F.3](#)), UMA timestamps (*infra* [§ III.G.2.a](#)), PPID hashing and encryption (*supra* [§ III.A.2](#)), and Google's data deletion practices (*infra* [§ III.C.2](#)).

⁵⁷ GOOG-BRWN-00029002, at -002-003.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

68. In my opinion, Google’s logging practices and server-side architecture for the Data at Issue, which is keyed by a pseudonymous identifier and stored in an orphaned state, renders Mr. Hochman’s proposal for identifying purported class members technically and practically impossible. *See infra* [§ III.G](#).

B. Opinion 2: Mr. Hochman’s Opinions On Interception, Notice, And Deletion Of Private Browsing Information (# 4, 5, 6, 26, 31) Are Contrary To Industry Guidelines On Private Browsing

69. Mr. Hochman contends that Google could have (i) redesigned Chrome or Google’s “tracking beacons” to “limit Google’s collection of private browsing information” (Hochman # 4) and (ii) notified users, at the time of collection, that it was collecting private browsing information (Hochman # 5). He opines that Google could have notified websites when it received private browsing information (Hochman # 6). Mr. Hochman asserts that Google “uniformly attempted to intercept all private browsing communications with non-Google websites that have a Google tracking beacon” (Hochman # 26). Finally, Mr. Hochman claims that “Google could delete from its systems records of Chrome Incognito browsing communications” (Hochman # 31). In my opinion, these five opinions are not credible because Mr. Hochman failed to consider the impact of the industry guideline that a website should not be able to detect whether a website visitor is in private browsing mode.

70. In my opinion, Mr. Hochman’s opinions on what Google could or should have done regarding private browsing (# 4, 5, 6, 26, 31) are marred by his failure to consider whether Apple, Firefox, Google and Edge are and should continue to intentionally take steps—in compliance with W3C TAG guidelines—to ensure that (i) the web remains accessible in private and normal browsing modes and (ii) the use of private browsing mode is not detectable by websites.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

71. Although Mr. Hochman refers to a number of different standards issued by Standard-Setting Organizations (“SSOs”),⁵⁸ he fails to acknowledge the industry guidelines most relevant to the allegations at issue: The W3C Technical Architecture Group (“TAG”) Observations on Private Browsing Mode. The World Wide Web Consortium (W3C) is the main international standards organization for the World Wide Web (WWW).⁵⁹ It was founded in 1994 by Tim Berners-Lee, the inventor of the WWW. The Technical Architecture Group (TAG) of W3C⁶⁰ consists of full-time W3C staff and representatives from industry leaders in web browsers and WWW, including Apple, Google, Microsoft, and Samsung.

72. In particular, Mr. Hochman fails to address the W3C TAG Observations on Private Browsing Modes published on July 5, 2019, which set forth the following guidelines on private browsing mode:

- “The Web should be accessible in private and normal browsing modes[.]”⁶¹
- “[T]he use of private browsing mode should not be detectable by websites[.]”⁶²

⁵⁸ See, e.g., Hochman ¶ 49 (“Google, other search engines, and website publishers have together adopted the Robot Exclusion Standard that prevents data from being harvested from websites by search engine robots, such as Google’s Googlebot, when a publisher opts out.”); ¶ 50 (“A specific example of Google failing to work with other browser vendors to establish privacy standards is the universal Do Not Track (‘DNT’) signal, which was first proposed in 2009.”).

⁵⁹ W3C, <https://www.w3.org> (last visited June 3, 2022).

⁶⁰ W3C, “W3C Tag,” <https://tag.w3.org> (last visited June 3, 2022).

⁶¹ W3C, “W3C TAG Observations on Private Browsing Modes,” W3C TAG Finding, <https://www.w3.org/2001/tag/doc/private-browsing-modes/> (July 5, 2019) (“When the differences in browser behavior between privacy and standard browsing modes can be detected because of standardization or implementation details, websites might choose to degrade browsing experience (for example, not displaying content) when they detect the users in private browsing modes. This is undesirable.”).

⁶² *Id.* (“[B]rowser vendors should work towards achieving private browsing mode work in a way indistinguishable [] from the normal mode, to respect the users’ privacy in choosing it.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

73. Mr. Hochman writes that Google designed Incognito mode on Chrome “so that no system like ads or Google Analytics can know whether or not you’re in incognito mode,”⁶³ but he fails to recognize that Google’s practice is consistent with the design principle set forth in the W3C TAG guidelines. In my opinion, providing a notice to websites at the time of collection (Hochman # 6) or redesigning Chrome or Google scripts (Mr. Hochman’s “tracking beacons”) to “limit Google’s collection of private browsing information” (Hochman # 4) would violate the W3C TAG principle that “the use of private browsing mode should not be detectable by websites[.]”⁶⁴

74. Similarly, Mr. Hochman appears to either have ignored or not have been aware of the W3C TAG guidelines when he suggested in Opinion 5 that “Google could have included either a pop-up notification or a choice to users at the time [of] collection [of information by Google services on third party websites], but it did not.”⁶⁵ As documents Google produced show, configuring Google scripts on third party websites to provide such a notification only to users in private browsing modes would require either (i) configuring the browser to send information flagging the user’s use of private browsing mode;⁶⁶ or (ii) configuring the embedded Google services to detect the use of Incognito mode.⁶⁷ In my opinion, both of these options would violate the W3C TAG guidelines discussed above and facilitate websites detecting the use of private browsing mode.

⁶³ Hochman ¶ 117 (quoting Berntson June 16, 2021 Tr. 282:11-15 (“Incognito mode was designed so that no system like ads or Google Analytics can know whether or not you’re in incognito mode . . .”)).

⁶⁴ W3C, “W3C TAG Observations on Private Browsing Modes,” W3C TAG Finding, <https://www.w3.org/2001/tag/doc/private-browsing-modes/> (July 5, 2019).

⁶⁵ Hochman ¶ 135.

⁶⁶ GOOG-BRWN-00554317, at -319; Adhya Nov. 19, 2021 Tr. 185:9-185:10; McClelland Feb. 18, 2022 Tr. 255:14-257:16.

⁶⁷ GOOG-BRWN-00554317, at -319; GOOG-CABR-03646925, at -927; Berntson, Mar. 18, 2022 Tr. 141:5-143:14.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

75. Third-party discovery indicates that web browser developers including Microsoft, Mozilla, and Apple are operating private browsing modes on Edge, Firefox, and Safari to comply with this W3C TAG guideline. In response to a subpoena Plaintiffs issued in this case requesting “ALL DOCUMENTS sufficient to IDENTIFY all individuals who have used InPrivate Browsing mode in Microsoft IE/Edge from June 1, 2016 to the present,” Microsoft responded that it “does not have documents responsive to the Request.”⁶⁸ In response to a subpoena Plaintiffs issued requesting “ALL DOCUMENTS sufficient to IDENTIFY all individuals who have used a Private Window or Private Browsing mode in Firefox from June 1, 2016 to the present,” Mozilla responded that it “does not possess documents or information sufficient to ascertain the identity of individuals who have used a Private Window or Private Browsing mode in Firefox.”⁶⁹ And in response to a subpoena seeking “ALL DOCUMENTS sufficient to IDENTIFY all individuals who have used a Private Window or Private Browsing mode in Safari,” Apple responded that it is “not in possession of the requested records.”⁷⁰

76. Mr. Hochman also asserts that Google uniformly attempted to intercept and collect data from all private browsing communications (Hochman # 26),⁷¹ and seems to suggest that Google could have avoided collecting data from private browsing communications. But browser vendors, following the W3C TAG guidelines, do not send an explicit signal from a browser instance to websites to detect whether a visitor is in private browsing mode. When Google receives the Data At Issue through its web-services (*i.e.*, Analytics, Ad Manager) from a Safari or Edge browser, then Google does not know whether the data it received came from a

⁶⁸ Microsoft’s Response to *Brown* Plaintiffs’ Subpoena (Aug. 27, 2021), at 4.

⁶⁹ Mozilla’s Response to *Brown* Plaintiffs’ Subpoena (Aug. 27, 2021), at 4.

⁷⁰ Apple’s Response to *Brown* Plaintiffs’ Subpoena (Sept. 20, 2021).

⁷¹ Hochman ¶ 308.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

browser in private browsing mode.⁷² Similarly, when Google receives the Data At Issue through its web-services from a Chrome browser the data does not arrive with an explicit signal designating it as coming from a browser in private browsing mode.⁷³

C. Opinion 3: Mr. Hochman’s Opinions On “Private Browsing Profiles,” Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate

77. I disagree with Mr. Hochman’s opinions that: (1) “Google, throughout the class period, created detailed profiles ... based on the private browsing information it collected” (Hochman # 10);⁷⁴ (2) private browsing information could be linked to information tied to the same user’s Google account (Hochman # 18) and the same user’s account with non-Google websites (Hochman # 19), but for Google’s deletion of the data (Hochman # 20).

1. Orphaned And Unidentified Interest Segments Are Not Cradle-To-Grave Profiles

78. I understand that Plaintiffs in this case have alleged that “[b]y tracking, collecting and intercepting users’ (including Plaintiffs’ and Class members’) personal communications indiscriminately—regardless of whether users attempted to avoid such tracking pursuant to Google’s instructions—Google has gained a complete, cradle-to-grave profile of users.”⁷⁵ This is incorrect. “Cradle-to-grave” means “from the beginning to end of life.”⁷⁶ Given the brief and isolated nature of the Data At Issue, it clearly does not meet this definition. Indeed, Mr. Hochman’s report concedes, at various points, that Google does not in fact combine data from regular browsing with the Data At Issue to create the “cradle-to-grave” profile Plaintiffs

⁷² See *infra* § III.H. for my rebuttal regarding Mr. Hochman’s opinion that the maybe_Chrome_incognito boolean field reliably detects Incognito mode.

⁷³ *Id.*

⁷⁴ Hochman ¶ 11.

⁷⁵ Complaint ¶ 93.

⁷⁶ Merriam-Websters.com Dictionary, ““From (the) cradle to (the) grave.” <https://www.merriam-webster.com/dictionary/from%20%28the%29%20cradle%20to%20%28the%29%20grave> (last visited June 6, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

have alleged.⁷⁷ Mr. Hochman does not show that the information contained in these profiles, which he concedes are limited to discrete Incognito browsing sessions, is combined with a user's signed-in or signed-out profile consisting of information from browsing that takes place outside of Incognito mode (or outside of other browsers' private browsing modes).⁷⁸ Notably, Mr. Hochman has used Chrome's Incognito mode in the past in other litigation to highlight this profile separation, while criticizing other experts for failing to do so.⁷⁹

79. Mr. Hochman cites to DBL [REDACTED] data associated with Biscotti IDs extracted from IDE cookies or mapped from PPIDs provided by the named Plaintiffs to claim that Google maintains detailed user profiles.⁸⁰ But as the DBL [REDACTED] data shows, the purported "profiles" are inferred interest segments keyed to pseudonymous identifiers unique to each discrete private browsing session or unique to the website publisher. Mr. Hochman opines that "[t]here may be several Google user profiles associated with a user, including GAIA-keyed user data and Biscotti-keyed user data."⁸¹ It is true that there may be multiple so-called user profiles

⁷⁷ See, e.g., Hochman ¶ 181 ("during the time-span of one incognito session personalization can happen based on the activities happening in that incognito session" (quoting Mardini Nov. 24, 2021 Tr. [294:4-8])); ¶ 187 ("[F]or purposes of Chrome Incognito, cookies such as Biscotti are specific to a particular session").

⁷⁸ See, e.g., Hochman ¶ 176 (stating that "[t]he stored user profile information from Incognito sessions contains similar content as user profile information from non-Incognito sessions based on my review of Plaintiffs' data" without contending that these profiles are linked or contain the same information).

⁷⁹ See, e.g., Declaration of Jonathan E. Hochman, *Telebrands Corp., v. Tinnus Enterprises, LLC*, Case No. PGR2015-00018 (USPTO 2015) ¶ 27 ("Amazon personalizes search results. In testing, I used the incognito feature of the Chrome browser to hide my identity from Amazon so that they would not personalize my search results based on my past search history."); Expert Report of Jonathan E. Hochman, *Rockwood Select Asset Fund XI, (6)-I, LLC v. Devine, Millimet & Branch, P.A.*, Case No. 1:14-cv-00303-LM, 2016 WL 4260622 (D.N.H. Jan. 29, 2016) at ¶ 8 ("Google personalizes search results based on a user's location and past search history. Mr. Levine's report does not indicate that he took any steps (such as using a Chrome Incognito window) to limit the personalization of the search results that he found.").

⁸⁰ Hochman ¶¶ 174-75, Appendix H.

⁸¹ Hochman ¶ 169 ("There may be several Google user profiles associated with a user, including GAIA-keyed user data and Biscotti-keyed user data.").

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

associated with the browsing activities of one user, but Google does not link these so-called profiles or re-identify any user from the pseudonymous private browsing data stored in an orphaned state. Simply put, it is my opinion that Google stores information keyed to pseudonymous identifiers unique to each private browsing session or unique to the website publisher subject to Google's data retention policy, and it does not maintain anything remotely resembling a cradle-to-grave profile of private browsing data for an individual user.

80. For example, Mr. Hochman states in Appendix H that “from the ‘2022-03-14 Brown v. Google - DBL [REDACTED] – AEO’ production, [REDACTED]’ contains Plaintiff Chasom Brown’s Incognito user profile information associated with Biscotti ID [REDACTED].”⁸² Mr. Hochman also notes that “this Biscotti ID corresponds to four Biscotti cookies,” and that “[w]hile the cookie values have changed over time, the embedded uid (Biscotti ID) has not.”⁸³ To illustrate, Mr. Hochman included the following table:⁸⁴

Item	Biscotti Cookie	uid	Creation Timestamp
1	[REDACTED]	[REDACTED]	[REDACTED]
2	[REDACTED]	[REDACTED]	[REDACTED]
90	[REDACTED]	[REDACTED]	[REDACTED]
98	[REDACTED]	[REDACTED]	[REDACTED]

⁸² See Hochman Appendix H ¶ 2.

⁸³ *Id.* ¶ 4.

⁸⁴ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

81. Mr Hochman's opinion is wrong for several reasons. *First*, Mr. Hochman suggests, but conspicuously stops short of saying, that the Data is associated with different cookie values from different private browsing sessions. It is not. As Mr. Hochman's table shows, the creation timestamp for these cookies are exactly the same: [REDACTED] which can be translated to human-readable time January 31, 2022 8:43:14 AM PT.⁸⁵ These four IDE cookies are simply re-encrypted versions of the same Biscotti ID unique to this one private browsing session.⁸⁶ Google resets the cookie value to prevent third-parties from using the IDE cookie value for tracking purposes.⁸⁷

82. *Second*, Google also produced DBL [REDACTED] data associated with other Biscotti IDs extracted from other IDE cookies provided by Plaintiff Brown, *see, e.g.*, [REDACTED] but Mr. Hochman does not even attempt to show that Google or anyone could link these two datasets to the same user. The data proves the opposite. For example, based on the produced DBL [REDACTED] data stored in column [REDACTED] on the same day associated with the two different Biscotti IDs, the user associated with Biscotti ID [REDACTED] [REDACTED] while the user associated with Biscotti ID [REDACTED] [REDACTED]. In my opinion, one cannot infer from the data that these two data sets keyed to different Biscotti IDs are associated with the same user. *See* [Appendix G. "Profile" Data](#).

⁸⁵ Epoch Converter, "Epoch & Unix Timestamp Conversion Tools" <https://www.epochconverter.com/> (last visited June 6, 2022).

⁸⁶ GOOG-CABR-04206179 at -185.

⁸⁷ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

83. Further, the DBL [REDACTED] data was pulled using Biscotti IDs decrypted from IDE cookies provided to Google by Plaintiff Brown under the Special Master process, which were copied from his browser before he closed the private browsing session (when the cookies would be deleted from the browser). But for Plaintiffs' representation that these cookies were pulled from his browser while in an Incognito session, Google does not know the identity of the user, or whether the data is private browsing data.

2. Mr. Hochman's Claims Regarding Google's Retention Policies Show That These Policies Prevent The Creation Of Such Profiles

84. Mr. Hochman also states that "it is my opinion that Google's deletion of certain data ... hinders the process of linking (or joining) a user's private browsing information to that user's Google account, and in some cases may make it impossible to do so where it would have been possible but for Google's deletion of the data."⁸⁸ As discussed *infra* at [§ IV.C.4](#), Google's retention, deletion, and anonymization of certain private browsing data aligns with best practices for data retention as outlined in CSC No. 3.⁸⁹ In my opinion, retaining data in perpetuity in order to "[aid] the process of linking (or joining) a user's private browsing information to that user's Google account," as Mr. Hochman appears to propose, would (i) conflict with CSC No. 3's guidance that "[d]ata retention must include both minimum and maximum timelines";⁹⁰ and (ii) increase the risk of fingerprinting by a bad actor. It is also my opinion that Google's deletion and anonymization of certain private browsing data undermines any claim that Google maintains "cradle-to-grave" profiles that track a user's private browsing mode activity for all time because, in addition to the session-based limitations that

⁸⁸ Hochman ¶ 21.

⁸⁹ "CSC No. 3" refers to Critical Security Control No. 3 from Version 8 of the Center for Internet Security's ("CIS") Critical Security Controls, which are recommendations for network security best practices issued by CIS. *See infra* [§ IV.C.4](#).

⁹⁰ Center for Internet Security, "CIS Critical Security Controls Version 8" <https://www.cisecurity.org/controls/v8> (last visited June 3, 2022), at -16.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Mr. Hochman concedes, any tracking is expressly time-limited (rather than indefinite) by virtue of Google's standard retention periods.

85. Based on my experience in the industry and the documents and testimony I have reviewed in this case, it is my opinion that the "private browsing profiles" on which Mr. Hochman opines are unidentified and orphaned, and not "cradle-to-grave" profiles.

D. Opinion 4: Mr. Hochman's Assertion That Google Used Private Browsing Information To Measure Conversions (# 14) Is Misleading

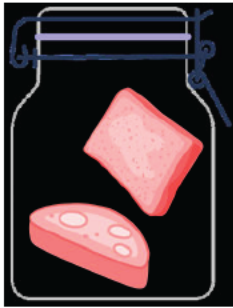

86. Mr. Hochman contends that "Google, throughout the class period, used private browsing information to measure and model conversions" (Hochman # 14).⁹¹ In particular, he asserts that "[t]hrough Google's server-side processes such as . . . [REDACTED] (linking a user's signed-out Biscotti IDs across different sessions) . . . Google creates 'a single-view of the user in all Display products' for ad targeting and measurement purposes."⁹²

87. In my opinion, these assertions are misleading because they create the impression that Google maps the Biscotti ID from one of the signed-out private browsing sessions at issue in this case to the Biscotti ID from a different signed-out private browsing session at issue in this case. Based on my review of documents and deposition testimony, there is no such mapping because the private browsing Biscotti IDs in this case are never in the same cookie jar as a GAIA ID.

⁹¹ Hochman ¶ 15.

⁹² Hochman ¶ 184 (citing GOOG-CABR-03652751).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

No Mapping / Measurement	Mapping / Measurement Could Occur
 <ul style="list-style-type: none"> • Biscotti ID # 2 (2483711983363751938) • Zwieback ID # 2 (0x43b763007d1ff50d) 	 <ul style="list-style-type: none"> • GAIA ID for USER@gmail.com • Biscotti ID # 1 (2515782358150873158)

88. Conversions are measured through mapping a GAIA ID to a Biscotti ID.

GOOG-CABR-03662096 at -100.

89. As Google employees have explained in sworn testimony, there is no such mapping for the Data at Issue because the relevant Biscotti cookies are never in the same cookie jar as the GAIA cookie.⁹³ In other words, it does not affect putative Class members.

⁹³ Berntson June 16, 2021 Tr. 372:14-273:10 (“Q. And I believe you testified that mapping in [REDACTED] requires a Gaia ID; is that -- is that right? A. That is correct. Q. For the dataflow that’s at issue in this case where users are on their browsers, they’re signed out of their Google accounts, they’re in private browsing mode, would there be any mapping from Gaia to Biscotti? A. No, because when you go into private browsing mode, you start off with a completely empty cookie jar. A Biscotti is created, and if you don’t sign in to Google, there’s no Gaia to map that new Biscotti to. The Biscotti that is present on the non-incognito browser instance is not shared

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

E. Opinion 5: Mr. Hochman's Description Of Entropy (# 18) Is Incorrect

90. Mr. Hochman describes entropy as a “metric for user identifiability . . . measured in number of bits of data needed to uniquely identify a person.”⁹⁴ He states that “[w]ith around 330 million people in the United States, 29 bits of data is more than sufficient to identify a person ($2^{29} = 537$ million).”⁹⁵ He states that “IP address and User Agent ‘carries 29.8 bits’ of entropy, which is more than sufficient to uniquely identify individuals in the United States.”⁹⁶

91. In my opinion, Mr. Hochman's description is incorrect and shows a profound misunderstanding of information entropy. Computing entropy bits and interpreting them correctly requires familiarity with probability and information theory.⁹⁷ To support my opinions, I have provided a detailed technical description of the relevant theory and formulas in [Appendix E](#). For readers without a background in probability and information theory, I explain the primary errors in Mr. Hochman's approach using general terms below.

92. **First**, contrary to Mr. Hochman's description, entropy is not a “metric for user identifiability.”⁹⁸ Entropy is a measure of the uncertainty of the outcome of a process.⁹⁹ Entropy

with the incognito browser instance so there's no way of creating that mapping from an incognito session.”).

⁹⁴ Hochman ¶ 231.

⁹⁵ *Id.*

⁹⁶ Hochman ¶ 233 (citing GOOG-CABR-04635379 at -380); *see also* Hochman ¶ 231 (asserting that “Google engineers measure that IP address alone contributes to 26.5 bits of entropy” (citing GOOG-BRWN-00601937)).

⁹⁷ Patrick Billingsley, “Probability and Measure,” Wiley, (3rd ed. 1995); Sheldon Ross, “Introduction to Probability Models,” Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014); T.M. Cover and J.A. Thomas, “Elements of Information Theory,” Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006).

⁹⁸ Hochman ¶ 231.

⁹⁹ For a detailed technical explanation, *see* [Appendix E. Entropy Bits and Fingerprint-ability: Formal Exposition](#), § [V.E.2](#); *see also* T.M. Cover and J.A. Thomas, “Elements of Information Theory,” Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

does not—and cannot be used to—establish what specific information is necessary to uniquely identify any specific person, or how many bits of data are “more than sufficient to identify a person.”¹⁰⁰ See *infra* § [V.E.2-3](#).

93. **Second**, whether IP address and User Agent (IP + UA) carry 29.8 bits, or 20 bits, or 40 bits has no bearing on user identification in this case because there is no reliable mapping relating IP + UA to an individual.¹⁰¹ Because an IP + UA pair: (i) may correspond to multiple devices and change dynamically, see *infra* § [III.G.1](#), and (ii) may be shared by multiple users, see *infra* § [III.G.1](#) and § [III.J](#), there is no reliable mapping from an IP + UA pair to an individual. Therefore, an IP + UA pair will not reliably identify the person who is using the browser regardless of the entropy bits of the identifier. Mr. Hochman explained this basic relationship of identifier to an individual in a recent publication:

*“An identifier by itself is meaningless and is just a code. For example, any random combination of nine numbers very well may be a social security number, but without identifying information, there is no relevance, utility or vulnerability. . . . Identifying information alone is not overly relevant, because it simply notes the existence of a person, without any detail of that person . . . sensitive information that cannot be linked to a specific person poses no risk to privacy and is the principal [sic] that allows large databases to exist for medical research.”*¹⁰²

¹⁰⁰ Hochman ¶ 231.

¹⁰¹ See [Appendix E](#), § [V.E.3](#) for a discussion on the necessity of a reliable mapping and § [V.E.7](#) for an example.

¹⁰² M. Fischer, J. Hochman, and D. Boffa, “Privacy-Preserving Data Sharing for Medical Research,” International Symposium on Stabilization, Safety, and Security of Distributed Systems, <https://cpsc.yale.edu/sites/default/files/files/TR1558.pdf> (Nov. 17–20, 2021) (“Hochman Paper”) at 2-3 (emphasis added).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

94. *Third*, entropy is an average-case metric.¹⁰³ It can provide insight into a system but not individuals. This limitation of entropy was explained by Martin Thomson, Distinguished Engineer at Mozilla, in a recent post:

“Information entropy remains useful as a means of understanding the overall utility of the information that a system provides. Understanding key statistics as part of a design is valuable. However, for entropy measures in particular, this is only useful from a perspective that seeks to reduce overall utility; *entropy provides almost no information about the experience of individuals.*”¹⁰⁴

Think of information entropy as analogous to average household income in the United States. Just like the single metric of average household income does not provide useful information regarding the actual finances of any specific household, entropy can provide insight to a system but not into any individual experience in it.

F. Opinion 6: Mr. Hochman’s Assertions On Fingerprinting (# 9, 18) Are Misleading And Unfounded

95. In his Opinion 18, Mr. Hochman opines “that Google throughout the class period and across the two classes *systematically collected and stored detailed private browsing data* that constitutes *what is commonly understood to be sensitive fingerprinting data*, which can be used to identify users and join data.”¹⁰⁵ Mr. Hochman does not describe a method to distinguish “fingerprinting data” from “sensitive fingerprinting data” and, in my experience, there is no common understanding of what fingerprinting data is “sensitive.” In his Opinion 9, Mr. Hochman claims that “‘pseudonymous’ *cookies such as Zwieback and Biscotti* cookies for

¹⁰³ Specifically, it is calculated as the average information conveyed by an outcome, averaged over all possible outcomes by weighting the information conveyed by an outcome with the likelihood of this outcome, see [Appendix E](#), Equation (2) in § [V.E.2](#) and associated discussion in § [V.E.3](#).

¹⁰⁴ “Entropy and Privacy Analysis,” Low Entropy, <https://lowentropy.net/posts/entropy-privacy/> (last visited June 3, 2022) (emphasis added).

¹⁰⁵ Hochman ¶ 223 (emphasis added).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

signed-out mode ... in fact, *can be linked to users through fingerprinting information Google collects in its logs, such as IP address, user agents, websites visited, and more*).”¹⁰⁶

96. In my opinion, Mr. Hochman’s assertions on fingerprinting¹⁰⁷ are misleading for the following reasons:

- ❖ **First**, Mr. Hochman does not directly address whether Google is engaged in fingerprinting thereby leaving the reader with the impression that Google may be engaged in fingerprinting. *See infra* § [III.F.1](#).
- ❖ **Second**, Mr. Hochman fails to acknowledge that Google’s internal policies, including policies cited by Mr. Hochman, expressly prohibit fingerprinting. *See infra* § [III.F.2](#).
- ❖ **Third**, Mr. Hochman fails to discuss or take into account the technical barriers that prevent fingerprinting. *See infra* § [III.F.3](#).

1. Google Does Not Engage In Fingerprinting

97. In the context of browser communications, fingerprinting refers to the combination of various bits of information to probabilistically identify a browser. Probabilistic identification of a browser through fingerprinting is different from the deterministic

¹⁰⁶ Hochman ¶ 160 (emphasis added).

¹⁰⁷ In the context of browser communications, fingerprinting refers to the combination of various bits of information to probabilistically identify a browser. Probabilistic identification of a browser through fingerprinting is different from the deterministic identification of an individual through an account log-in. Actors attempting fingerprinting most commonly use network-related information such as an IP address and web browser-related information such as the type and version of the browser (*i.e.*, User Agent). Other popular sources for fingerprinting attempts include presentation-related information (font sizes and screen resolution), hardware-related information (GPU characteristics), and applications-related information (installed browser extensions). *See* P. Laperdrix, W. Rudametkin, and B. Baudry, “Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints,” IEEE Symposium on Security and Privacy, San Jose CA, <https://ieeexplore.ieee.org/abstract/document/7546540> (2016); Y. Cao, S. Li, and E. Wijmans, “(Cross-)Browser Fingerprinting via OS and Hardware Level Features,” NDSS, San Diego CA, https://www.ndss-symposium.org/wp-content/uploads/2017/09/ndss2017_02B-3_Cao_paper.pdf (2017).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

identification of an individual through an account log-in. Actors attempting fingerprinting most commonly use network-related information such as an IP address and web browser-related information such as the type and version of the browser (i.e. User Agent). Other popular sources for fingerprinting attempts include presentation-related information (font sizes and screen resolution), hardware-related information (GPU characteristics), and applications-related information (installed browser extensions).¹⁰⁸

98. In my opinion, Mr. Hochman's statements on fingerprinting are misleading because they create the impression that Google may be engaged in fingerprinting to identify users. Plaintiffs allege that "Google also builds its profile of users (including Plaintiffs and Class members) by 'fingerprinting' techniques."¹⁰⁹ Plaintiffs also allege that "Google accomplishes its surreptitious interception and data collection through means that include Google Analytics, Google 'fingerprinting' techniques, concurrent Google applications and processes on a consumer's device, and Google's Ad Manager."¹¹⁰ Mr. Hochman's report does not address Plaintiffs' bold and completely unsupported assertions of fingerprinting.

99. Mr. Hochman does not show that Google actually engages in fingerprinting. Instead, he merely states that "Google *receives* a myriad of fingerprinting information."¹¹¹ There is a profound difference between engaging in fingerprinting and "receiv[ing] . . . fingerprinting information." The former requires a recipient of "fingerprinting information" to use such

¹⁰⁸ See P. Laperdrix, W. Rudametkin, and B. Baudry, "Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints," IEEE Symposium on Security and Privacy, San Jose CA, <https://ieeexplore.ieee.org/abstract/document/7546540> (2016); Y. Cao, S. Li, and E. Wijmans, "(Cross-)Browser Fingerprinting via OS and Hardware Level Features," NDSS, San Diego CA, https://www.ndss-symposium.org/wp-content/uploads/2017/09/ndss2017_02B-3_Cao_paper.pdf (2017).

¹⁰⁹ Complaint ¶ 100.

¹¹⁰ *Id.* ¶ 8.

¹¹¹ Hochman ¶ 107 (emphasis added).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

information to attempt to probabilistically identify a particular user or device. In my opinion, Google does not engage in fingerprinting as Plaintiffs allege in their complaint,¹¹² and Mr. Hochman's failure to directly address whether Google engages in fingerprinting is misleading.

2. Google's Internal Policies Expressly Prohibit Fingerprinting

100. Mr. Hochman fails to acknowledge that Google has long-standing internal policies that expressly prohibit Googlers from engaging in (i) the fingerprinting that Plaintiffs alleged and (ii) the fingerprinting that Mr. Hochman is proposing Google engages in an attempt to identify class members. *See infra* [§ III.G.1](#).

101. Google has a number of policies and guidelines concerning the collection, storage, usage and deletion of data. These policies and guidelines include, for example:

- the Device/App/Browser Fingerprinting and Immutable Identifiers Policy;¹¹³
- the User Data Access Policy;¹¹⁴ and
- the User Data Retention and Deletion Policy.¹¹⁵

102. Google's internal policies expressly prohibit fingerprinting unless it is to prevent spam, abuse, fraud, or other such user-beneficial usages completely unrelated to Plaintiffs'

¹¹² Complaint ¶¶ 8, 100.

¹¹³ GOOG-BRWN-00029326.

¹¹⁴ GOOG-CABR-05455683.

¹¹⁵ GOOG-CABR-00073922.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

allegations.¹¹⁶ Additionally, Google identified requests for exceptions to Google’s fingerprinting policies described in Google’s Initial and Supplemental Responses and Objections to Plaintiffs’ Interrogatory No. 17.¹¹⁷ These exceptions are completely unrelated to Plaintiffs’ allegations that Google “builds its profile of users (including Plaintiffs and Class members) by ‘fingerprinting’ techniques.”¹¹⁸ Last, Google’s strict enforcement of this policy was confirmed by every deposed Google employee or former employee who Plaintiffs asked about Google’s fingerprinting policy.¹¹⁹

¹¹⁶ See GOOG-CABR-04720562, at -563; GOOG-CABR-00073873, at -873

[REDACTED] GOOG-BRWN-00029433, at -435

[REDACTED] GOOG-CABR-00086797, at -797

¹¹⁷ See Google’s Supplemental Objections and Responses to Plaintiffs’ Interrogatories (No. 17), at 4 (“In the context of web browsing on non-Google websites, exceptions to Google’s anti-fingerprinting policy have been requested for (i) analyzing entropy in Javascript APIs as part of ‘efforts at throttling browser fingerprinting on the web’ . . . ; and (ii) analyzing certain ad blocker features.”).

¹¹⁸ Complaint ¶ 100.

¹¹⁹ See, e.g., McClelland Tr. 279:3-11 (“Q. Google also prohibits fingerprinting users for the purpose of associating their activity over time or across contexts, is that right? . . . A. Yes, that is my understanding, that fingerprinting was also not allowed to be used.”); *id.* at 303:10-304:3 (“Q. Based on your work as product lead for Chrome browser privacy, would you agree that Google identifies users with fingerprinting techniques? A. I know that it was not allowed by policy and I never saw any evidence of it happening either. Q. Does Google use fingerprinting to identify users and personalize advertising, to your knowledge? A. I believe there may be some legitimate uses for fingerprinting around anti-fraud with sign-in, but understanding, again, is that it’s not used for ad targeting and, again, I never saw any evidence to counter that.”); Adkins Apr. 14, 2021 Tr. 188:5-10 (“[T]he Google services do not use any combination of identifiers to try to uniquely identify users, other than the Google logged in cookie and so, therefore, it’s impossible because we don’t conduct -- that’s against our privacy policy.”); *id.* at 314:2-8 (“[N]ot only is the practice discouraged or forbidden, but I am aware of there are active measures taken to prevent teams from being able to do fingerprinting within Google, so that it’s not accidentally done, so fingerprinting is not accidentally done.”); Bindra Feb. 8, 2022 Tr. 123:1-3 (“Q. So Google does not engage in fingerprinting; is that what you’re saying? A. Correct.”); Halavati Jan. 18, 2022 Tr. 151:2-5 (“Google has . . . strong regulations against

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

3. Google's System Architecture Supports Google's Anti-Fingerprinting Policy

103. Google has implemented a server architecture referred to as the [REDACTED],¹²⁰ which is a single service that handles all raw user identity and privacy signals (in order to, *inter alia*, reduce the risk of re-identification of pseudonymous data¹²¹), enabling the rest of the Display Ads stack to become user identifier-agnostic.

104. Because only [REDACTED] has access to sensitive information that is required to engage in the fingerprinting that Plaintiffs allege in their complaint, this architecture reduces the risk of mistaken violations of the anti-fingerprinting policy, and makes it hard for a bad actor (*i.e.*, a rogue employee) to fingerprint.

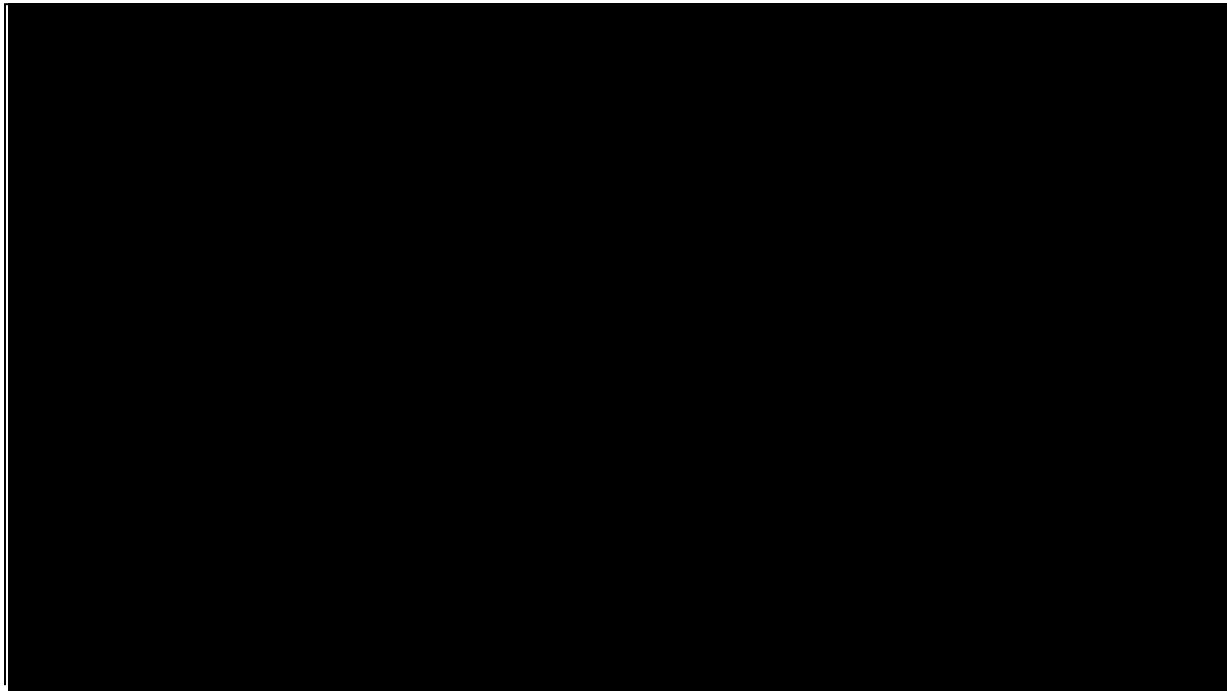
fingerprinting and all -- all user identifiable data that is collected should be by direct user consent or signing into a Google account.”); Monsees Apr. 9, 2021 Tr. 314:23-315:3 (“Q. So why is it Google’s policy that you must not fingerprint users for the purpose of associating a user’s activity over time or across contexts tracking? A. That would violate our policies and, I think, statements to our users and regulators.”); Shelton Mar. 2, 2022 Tr. 141:10-16 (“Q. When you were working at Google, to your knowledge, was Google engaged in fingerprinting? . . . [A.] I’m not aware of Google doing what I have characterized here as fingerprinting.”).

¹²⁰ GOOG-BRWN-00160342; GOOG-CABR-04715843.

¹²¹ GOOG-CABR-00058557, at -585-586; GOOG-CABR-00058926, at -963.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

105. [REDACTED] uses encryption to ensure that GAIA and Biscotti-keyed data is segregated.¹²²



106. The encryption keys for GAIA and Biscottis stored in these logs are stored separately, which reduces the risk of re-identification of unauthenticated data.¹²³

107. [REDACTED] processing takes place “offline,” which refers to automated processes that occur without direct human involvement.¹²⁴ This further reduces the risk of joining of signed-in

¹²² GOOG-CABR-05466323, at -333; *see also* Berntson June 16, 2021 Tr. 121:13-122:9 (“[W]hat [REDACTED] does is it extracts all user identifiers, device identifiers, et cetera, and it partitions them, so that within a given ad request you -- you really are only going to be able to use one identifier, and all the downstream systems have sort of additional combinations of identifiers made unavailable. And we designed this system because we have a number of internal policies at Google that say you can’t reidentify users, you can’t combine logged-in users and non-logged-in users, and that goes with what we call a data minimization philosophy that we design our systems very explicitly to make available only the information that’s necessary to run the business, because at the end of the day, we believe it’s important to protect user privacy.”).

¹²³ Berntson June 16, 2021 Tr. 121:10-122:9.

¹²⁴ *Id.* at 126:4-11; GOOG-CABR-03841628; GOOG-CABR-04721001.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

and signed-out data by, for example, preventing engineers working on [REDACTED] from directly accessing the data that [REDACTED] processes.

108. Based on my experience in the industry and the reports, documents and testimony I have reviewed in this case, I have not seen *any* evidence of Google engaging in fingerprinting. In my opinion, Google's policies prohibit and practices prevent it from engaging in the fingerprinting Plaintiffs alleged and Mr. Hochman is proposing Google use to identify Class I. My conclusion is consistent with Google's anti-fingerprinting policies and testimony from Google employees (discussed above at ¶¶ 100-102). Certainly Mr. Hochman has not pointed to any evidence that Google engages in fingerprinting.

G. Opinion 7: Mr. Hochman's Proposal To Identify Class I (Chrome Class) (# 22) Is Unreasonable and Unreliable

109. Mr. Hochman's proposal to identify class members for Class I (Chrome Class)¹²⁵ consists of the following steps:

- a. Query Google's logs to "find instances of users browsing within Incognito on [a] non-Google website while signed out of Google."¹²⁶
- b. Extract information (IP Address, User Agent, PPID, Analytics User ID, UMA ID) from log data associated with a user browsing within Incognito on a non-Google website.¹²⁷
- c. Run the extracted information (IP Address, User Agent, PPID, Analytics User ID, UMA ID) over logs that are written when a user is signed in to Google.¹²⁸

¹²⁵ Complaint ¶ 192 ("All Chrome browser users with a Google account who accessed a non-Google website containing Google tracking or advertising code using such a browser and who were (a) in 'Incognito mode' on that browser and (b) were not logged into their Google account on that browser, but whose communications, including identifying information and online browsing history, Google nevertheless intercepted, received, or collected from June 1, 2016 through the present (the 'Class Period').").

¹²⁶ Hochman ¶ 288.

¹²⁷ Hochman ¶ 293.

¹²⁸ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- d. If the extracted information (IP Address, User Agent, PPID, Analytics User ID, UMA ID) from step b exists in logs associated with Google Accounts, then pull the associated account information and include the account holder in the class.¹²⁹

110. In my opinion, this would be a highly unreliable method for identifying class members and would result in widespread mis-identification for the following reasons:

- ❖ **First**, As I explain in my Opinion 2 there is no explicit signal sent from the Chrome browser to Google’s web-services to detect Incognito traffic, *see supra* [§ III.B](#). As I explain in my Opinion 8, Google’s web-services can not reliably detect Incognito traffic in Google’s logs due to false positives associated with the “maybe_chrome_incognito” bit. *See infra* [§ III.H](#). Therefore, since the first step in Mr. Hochman’s proposal to identify Class I requires the accurate detection of Incognito traffic, and Google can not reliably detect Incognito traffic, his method will not accurately identify class members.
- ❖ **Second**, while combining IP address + User Agent can, in some cases, result in sufficiently unique information to distinguish between distinct devices, it cannot be used to reliably identify the specific device from which data was received or the individual that was browsing. And the process of trying to do so definitely cannot be replicated to identify millions of potential devices used in private browsing *and* the purported class members (composed of individuals) who did the private browsing. *See infra* [§ III.G.1](#).
- ❖ **Third**, the pseudonymous identifiers Mr. Hochman proposes for joining logged-out Incognito traffic with the user’s Google Account (*e.g.*, UMA ID, PPID, Analytics User ID, Biscotti ID) will not reliably identify class members. *See infra* [§ III.G.2](#).

¹²⁹ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

1. Hochman's IP + UA Fingerprinting Method Will Not Work

111. I disagree with Mr. Hochman's opinion that "an IP address, especially when combined with a user-agent string, constitutes personally identifiable information ('PII') because this data can be used to uniquely identify a user with a high probability of success."¹³⁰ The combination of IP address and User Agent that Mr. Hochman proposes for identifying class members via his proposed fingerprinting methodology is not sufficiently unique to identify class members because there are many situations where more than one user will have an identical IP address and user agent. Specifically, I will explain that neither the IP address, nor the User Agent, nor their combination is a reliable method to identify class members because there is no reliable one to one mapping from IP + UA to an individual class member.

a. IP Addresses Are Neither Unique Nor Static

112. An IPv4 address is a set of four numbers, each ranging from 0 to 255, assigned to an internet-connected device or devices. Each address consists of 32 bits, yielding a total of 2^{32} possible addresses. With the increase of the number of devices per person and the people using the Internet, these IP addresses were not enough and a newer version of the IP protocol, IPv6, defined new IP addresses each consisting of 128 bits. Adoption of IPv6 worldwide is still less than 40 percent and currently at 46.74 percent in US,¹³¹ and thus at least 50 percent of the IP addresses in the US are IPv4 addresses. Unless otherwise stated, I use the term IP address to refer to both IPv4 and IPv6 addresses.

113. It is rarely the case that a device has a unique, static IP address. For example, devices inside the home of a family of four who own a total of eight devices combined will all share the same external IP address, which is then internally translated to eight distinct local IP

¹³⁰ Hochman ¶ 105.

¹³¹ GoogleIPv6, <https://www.google.com/intl/en/ipv6/statistics.html> (last visited June 3, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

addresses using the so-called NAT (Network Address Translation) protocol.¹³² There is no way to tell which of the eight devices is the one that accesses a website, if all one has as an identifier is the external IP address of the device.

114. As another example, employees of a large company who are working from home are connected to their company via what is called a VPN (Virtual Private Network). For the vast majority of VPN services, if any of these employees access a website, the external IP address will be the same for each employee and will equal the IP address of the VPN server.¹³³ Again, there is no way to know which of the hundreds of potential devices connected to the VPN server is the one that accessed the website, if the only identifier is the external IP address.

115. As one more example, consider a company which assigns dynamic rather than static IP addresses to devices inside the company's network, using the so-called DHCP (Dynamic Host Configuration) Protocol.¹³⁴ Note that due to the flexibility of DHCP over static IP addresses, it is the preferred method to assign IP addresses to end-user devices across large organizations.¹³⁵ Identifying a device by its IP address when DHCP is used is very unreliable because one cannot know for sure which device has a specific dynamic IP among the possible set of IP addresses at any given time, since the assignment of IP addresses to devices changes dynamically. For example, in a company with 200 devices and 256 IP addresses, any device

¹³² J.F. Kurose & K.W. Ross, *Computer Networking: A Top-Down Approach*, Ch. 4 (8th ed. 2020).

¹³³ Some VPN service providers offer, for an additional cost, the option of a dedicated IP address. In this case, the client IP address will still be hidden, but the client's internet traffic will not show the shared VPN server IP address but rather this dedicated IP address. *See, e.g.*, NordVPN, <https://nordvpn.com/features/dedicated-ip/> (last visited June 3, 2022). This feature does not target institutional clients of VPN services like employers, but rather individual clients who want to avoid the so-called "bad neighbor" effect.

¹³⁴ J.F. Kurose & K.W. Ross, *Computer Networking: A Top-Down Approach* Ch. 4 (8th ed. 2020).

¹³⁵ Microsoft, "Dynamic Host Configuration Protocol (DHCP)," <https://perma.cc/7N5L-QKCO> (July 29, 2021); J.F. Kurose & K.W. Ross, *Computer Networking: A Top-Down Approach* Ch. 4 (8th ed. 2020).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

may have any of the 256 IP addresses at any given time. There is no way to know for sure which of the 200 devices is the one that accesses a website, if the only identifier is the external IP address of the device that accesses the website.

116. One more reason why a device may not have a unique, static IP address is the use of Onion Routing, known to most as Tor.¹³⁶ Tor directs internet traffic through an overlay network consisting of thousands of relays and uses a series of layered nodes to hide the IP address of a device. In practice, a device that uses Tor to access a website cannot be identified by the observed IP address, which is the IP address of the so-called exit node.

117. Because the number of possible IPv6 addresses is so vast, there is a misconception that an IPv6 address always identifies a single device. This is a gross misconception. First, VPN services that support IPv6, hide both IPv4 and IPv6 device addresses in a seamless fashion.¹³⁷ Second, IPv6 NAT products (*see*, for example, IPv6 NAT functionality within Junos OS of Juniper Networks) not only support address translation between IPv4 and IPv6 addresses, but also between IPv6 hosts. In particular, NAT between IPv6 hosts is done in a similar manner as IPv4 NAT, and, in this case, a single, public IPv6 address may correspond to a large number of private IPv6 addresses. In addition, to address privacy concerns with dynamic IPv6 addresses assigned by DHCPv6, privacy extensions allow clients to use random lower 64bit IIDs.¹³⁸ For the upper 64bit IID, providers employ prefix rotation via what is known as temporary mode DHCPv6,¹³⁹ made possible thanks to the sheer number of IPv6 addresses which

¹³⁶ Tor Project, <https://www.torproject.org/about/history/> (last visited June 3, 2022).

¹³⁷ NordVPN, <https://nordvpn.com/features/hide-ip/> (last visited June 3, 2022); Surfshark, <https://surfshark.com/use-cases> (last visited June 3, 2022).

¹³⁸ T. Narten, R. Draves, and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," RFC 4941 (Draft Standard), <https://www.rfc-editor.org/rfc/pdf/rfc4941.txt.pdf> (Sept. 2007).

¹³⁹ Tomek Mrugalski, et al., "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," RFC 8415 (Proposed Standard), <https://www.rfc-editor.org/rfc/pdf/rfc8415.txt.pdf> (Nov. 2018).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

has resulted in single residential customers often having more IPv6 addresses assigned than the entire IPv4 address space.¹⁴⁰ The combined result of the above random selection of IPv6 address bits, together with the large IPv6 address space, has resulted in many IPv6 addresses being used only once, *see*, for example, data from a large CDN¹⁴¹ which found that more than 90 percent of IPv6 addresses appear only once in a long-running data collection campaign.¹⁴² Last, Tor has recently added support for IPv6 addresses.¹⁴³

118. Additionally, Google applies [REDACTED]

[REDACTED]

[REDACTED].¹⁴⁴

¹⁴⁰ E. Rye, R. Beverly, and K. Claffy, “Follow the Scent: Defeating IPv6 Prefix Rotation Privacy,” Proceedings of ACM Internet Measurement Conference (IMC), <https://arxiv.org/pdf/2102.00542.pdf> (Nov. 2-4, 2021).

¹⁴¹ D. Plonka and A. Berger, “Temporal and Spatial Classification of Active IPv6 Addresses,” Proceedings of ACM Internet Measurement Conference (IMC), <https://www.akamai.com/site/en/documents/research-paper/temporal-and-spatial-classification-of-active-ipv6-addresses-technical-publication.pdf> (Oct. 28-30, 2015).

¹⁴² While the use of one-time IPv6 addresses is prevalent among non-mobile devices, cellular providers may assign static IPv6 addresses to their clients.

¹⁴³ The State of IPv6 support on the Tor Network, <https://blog.torproject.org/state-of-ipv6-support-tor-network/> (last visited June 3, 2022).

¹⁴⁴ GOOG-BRWN-00029002, at -002.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



119. Mr. Hochman does not account for this IP address redaction process. And in my opinion, it makes IP addresses even less suitable for joining authenticated and unauthenticated data via Mr. Hochman's proposed fingerprinting methodology because IP addresses are even less identifiable when the lower 8 bits from IPv4 and the lower 80 bits from IPv6 are removed.

120. That an IP address cannot reliably identify a device due to the widespread adoption of NAT, VPN and DHCP is also evident by analyzing data Google produced under the Special Master process. Out of the 4,945 distinct IP addresses, there are 159 that have multiple GAIA's associated with them, and three of these IP addresses, specifically 172.58.30.176, 72.58.27.240, and 172.58.110.200, have multiple GAIA's that correspond to *more than one* plaintiff.¹⁴⁵ In other words, Plaintiffs' own data which correspond to five different individuals who presumably do not share the same household (hence they would not live on the same hose behind a NAT) and presumably do not work for the same company (hence they would not be

¹⁴⁵ See [Appendix F, IP Address + User Agent Data Analysis](#).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

behind the same VPN or served by the same DHCP server), *show common IP addresses for two different Plaintiffs.*

b. User Agent Strings Are Neither Unique Nor Static

121. A User Agent string (UA) contains information about the type of the browser (e.g. Chrome, Edge, Mozilla, Safari), the version of the browser, and the operating system over which the browser is running (e.g. Windows, macOS, iOS, Linux).¹⁴⁶ It is used to identify the type and version of the browser and the operating system such that the behavior and content of web browsing can be customized accordingly. It should be evident even to a layman that millions of users share the same UA. What is more, it should be equally evident that some UAs are more common than others, for example, recent versions of popular web browsers running on top of popular operating systems are very common across devices. In fact, a recent study, which collected UAs from an Internet measurement company over the course of two years, found that the top 10 most popular UAs correspond to 26 percent of daily traffic.¹⁴⁷ With about 50 billion total HTTP requests per day in the collected data, this implies that more than one billion daily HTTP requests correspond to the same UA.¹⁴⁸ If a device makes on average 10-100 HTTP requests during a day, this would imply that tens of millions of devices from which data have been collected share the same UA.

122. The fact that a UA is usually shared by many devices is also evident by merely looking into the data submitted by the Plaintiffs. Out of the 1,237 unique UAs across all HTTP

¹⁴⁶ MDN Plus, “User-Agent,” <https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/User-Agent> (last visited June 3, 2022); Chrome Developers, “User-Agent Strings,” <https://developer.chrome.com/docs/multidevice/user-agent/> (updated Nov. 9, 2021).

¹⁴⁷ J. Kline, P. Barford, A. Cahn, and G. Sommers, “On the structure and characteristics of user agent strings”, ACM Sigcomm, <https://conferences.sigcomm.org/imc/2017/papers/imc17-final253.pdf> (Nov. 1-3, 2017).

¹⁴⁸ 26 percent of the 50B HTTP requests is more than 12.5B HTTP requests. Hence, on average, each of the 10 most popular UAs corresponds to more than 1.25B HTTP requests.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

requests of the five Plaintiffs, 113 of these UAs have multiple GAIAs associated with them, and, more telling, 73 of these UAs are *common* among multiple Plaintiffs. For example, three of these UAs are common among four out of the five Plaintiffs, namely “Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/95.0.4638.69 Safari/537.36,gzip(gfe)” and “Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/96.0.4664.45 Safari/537.36,gzip(gfe),” and “Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/98.0.4758.102 Safari/537.36,gzip(gfe),gzip(gfe).”¹⁴⁹ As another example from the data shared by the Plaintiffs, Mr. Hochman’s Appendix H contains 17 data entries corresponding to the five Plaintiffs, and even in this small data set of merely 17 entries of five people, three (60 percent) of them share the same UA.¹⁵⁰

123. Additionally, as Mr. Hochman has recognized in public statements, UA is not static and can be changed by the user via the use of, *e.g.*, a “user agent switcher” plugin.¹⁵¹ Mr. Hochman does not account for the ability of users to change their UA through the use of such tools. In situations where an IP address is shared by multiple users, the use of such tools could lead to false positives and render Mr. Hochman’s proposed class member identification methodology over-inclusive. Consider, for example, a shared Wifi connection where one user changes his or her user agent (the “Switcher”) to a user agent string that matches another user

¹⁴⁹ See [Appendix F. IP Address + User Agent Data Analysis](#).

¹⁵⁰ See Hochman Appendix H ¶¶ 13, 27, 31 (Plaintiffs Brown, Byatt, and Davis share the same UA: “Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/97.0.4692.99 Safari/537.36”).

¹⁵¹ Twitter, <https://twitter.com/Jehochman/status/1153277584542711808> (last visited June 3, 2022) (citing Lawrence Abrams, “How to Switch Back to the Old Twitter Layout,” Bleeping Computer, <https://www.bleepingcomputer.com/news/technology/how-to-switch-back-to-the-old-twitter-layout/> (July 16, 2019)) (“A method that continues to work is to change the browser's user agent string to the one used by Internet Explorer 11 when using Twitter. As Twitter's new layout does not support Internet Explorer 11, they switch you to [the old] layout.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

connected to the same Wifi who has never used private browsing mode (the “Regular Browsing User”). The IP address and user agent of the Switcher using private browsing mode would match the IP address and user agent of the Regular Browsing User, which could lead to the Regular Browsing User’s (incorrect) inclusion in the class because searching for the Switcher’s IP address and user agent will return results for the Regular Browsing User.

c. The Combination Of An IP Address And A User Agent String Is Neither Unique Nor Static

124. Mr. Hochman argues that the combination of an IP address and the UA may be used to identify class members. However, this method is unreliable. First, it is evident that devices which have the same IP address may also have the same UA. As a matter of fact, it is more likely for two devices to have the same UA when they have the same IP address, than when they do not. To see this, consider, for example, a virtual private network (VPN) used by a business organization for 100 remote workers.¹⁵² All devices connected to the VPN will show an identical external IP address. Because IT departments often maintain and upgrade company devices in batches, browser types and versions as well as operating systems will likely match

¹⁵² VPNs are also used by individual users for a number of reasons, including to specifically mask a device’s IP address in order to access content or services that are not available in the state or country where a device is physically located. *See, e.g.*, Arjun Ruparelia, “Best DraftKings Sportsbook VPN in 2022: Unblock DraftKings From Anywhere With a VPN,” Cloudwards, <https://www.cloudwards.net/draftkings-sportsbook-vpn/> (Apr. 27, 2022) (“Planning a quick trip to a state where DraftKings is inaccessible? You’ll need a DraftKings Sportsbook VPN to get your daily dose of sports betting. If you’re confused about which VPN to choose, we give you the five best ones that can access DraftKings from anywhere.”); Osman Husain, “How to watch ESPN anywhere with a VPN,” Comparitech, <https://www.comparitech.com/blog/vpn-privacy/best-vpn-espn/> (Jan. 19, 2022) (“Unfortunately, ESPN isn’t available everywhere. This can cause problems, for instance, if you’re trying to watch your favorite team play while on vacation abroad (outside of the US). Even assuming you can stream ESPN where you are, there’s a good chance that some sports are subject to regional restrictions (sometimes called blackouts) . . . On the plus side, it’s fairly easy to stream ESPN live online with the right VPN. Short for Virtual Private Network, a VPN connection masks your real location so that you can browse safely and regain access to your usual services while traveling.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

across company devices, resulting in the same UA. Therefore, assuming the organization provides its employees with identical laptops with Chrome (or any other browser) as the default browser (and automatic updates enabled), the user agent and IP address will be identical for all 100 employees' laptops.

125. Consider, as another example, a wifi connection on a commercial passenger jet. If the wifi is configured to use one external public IP address with NAT configured to translate all local IP addresses to that same public IP address, then everyone using the jet's wifi router will have the same IP address. If for example, there are 100 passengers connected to the same router, and ten of them are using the most recent version of the Safari browser on the same model of iPhone,¹⁵³ all ten of those users will have an identical user agent and IP address.

126. Because IT departments often maintain and upgrade company devices in batches, browser types and versions as well as operating systems will likely match across company devices, resulting in the same UA. With potentially thousands of devices sharing a single VPN IP address,¹⁵⁴ hundreds of millions of devices connecting to the Internet via a VPN services in

¹⁵³ This scenario is not implausible in light of (i) the popularity of certain models of handheld devices, *see* M. Levin and J. Lowitz, "iPhone 13 Models Have Biggest Share in Years," Consumer Intelligence Research Partners LLC, <https://files.constantcontact.com/150f9af2201/06eda1e6-cb00-4462-836f-73aa0120e439.pdf?rdr=true> (Apr. 21, 2022) ("Among all individual models, the iPhone 13 had the largest [market] share of any single model [in Q1 2022], with 38%, which was the largest share for any single model in many quarters. In the year-ago March 2021 quarter, the year-old iPhone 11 had the largest share, at 24%, while the iPhone 12, comparable to the current iPhone 13, had 22% of sales. The iPhone 13 share of 38% in this quarter nearly doubles last year's iPhone 12 share."); and (ii) broad enabling of automatic updates on such devices, *see* S. O'Dea, "App-updating habits of smartphone users in the United States 2016," Statista, <https://www.statista.com/statistics/747569/united-states-survey-smartphone-users-app-update-frequency/#statisticContainer> (Feb. 28, 2020) (2016 survey results showing 32 percent of users enabled automatic updates on their smartphone).

¹⁵⁴ There is no limit on how many connections may be served under the same VPN IP address, as VPN services use load-balancing among multiple servers clustered together to scale the number of supported connections. *See, e.g.,* Microsoft, "High Availability," <https://docs.microsoft.com/en-us/windows-server/remote/remote-access/vpn/always-on-vpn/dep/always-on-vpn-adv-options#high-availability> (last visited June 3, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

the United States,¹⁵⁵ and within the general population of users/devices, the 10 most popular UAs corresponding to at least a quarter of the total traffic/devices,¹⁵⁶ it is evident that the combination of an IP address and a UA is not unique since it may correspond to many different users. Moreover, since IP addresses are not static, for example, due to DHCP, and UAs are not static either, for example, because every browser update will change the UA, the combination of an IP address and a UA is not static either. For these reasons, the combination of an IP address and a UA cannot reliably identify devices, much less class members.

127. In sum, neither IPv4 nor IPv6 addresses can be used on their own right to consistently identify devices with high enough probability of success to reliably identify class members, and, for the reasons explained above, neither type of IP address in combination with UA may be used to reliably identify class members.

2. Pseudonymous IDs Can Not Be Used To Reliably Identify Individuals

128. The second step in Mr. Hochman's proposal is to extract information (IP Address, User Agent, PPID, Analytics User ID, UMA ID) from log data associated with a user browsing within Incognito on a non-Google website. I disagree with his opinion that these can be used to identify individuals, and I take each in turn below.

a. UMA ID

129. Mr. Hochman describes the process as follows for using UMA ID:

"Even if a user can no longer locate their UMA ID in an old device that they no longer possess or if the user factory resets a device, it is possible for Google to find a user's UMA ID by correlating the user's Gaia logs and UMA logs with search actions and timestamps. Knowing that detailed event level data are stored with timestamps, IP

¹⁵⁵ Aleksandar Kochovski, "The Top 25 VPN Statistics, Facts & Trends for 2022," <https://www.cloudwards.net/vpn-statistics/> (Mar. 18, 2022).

¹⁵⁶ J. Kline, P. Barford, A. Cahn, and G. Sommers, "On the structure and characteristics of user agent strings," ACM Sigcomm, <https://conferences.sigcomm.org/imc/2017/papers/imc17-final253.pdf> (Nov. 1-3, 2017).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

address and user agent in these logs, it is my opinion that UMA can be joined with ads event logs.”¹⁵⁷

130. In my opinion, UMA IDs cannot be used, either independently or in conjunction with other data, to identify individual users. As an initial matter, Google witnesses have testified that UMA data cannot be used independently to identify individual users.¹⁵⁸ This is so for a number of reasons:

- a. UMA data is only sent for users who have opted into Chrome metrics, and opt-in rates vary by device type. *See, e.g.*, GOOG-CABR-00057895, at -896:



- b. UMA data regarding Incognito usage is an approximation obtained via sampling (*i.e.*, an approximation of the total rate of Incognito usage extrapolated from a

¹⁵⁷ Hochman ¶ 258 (citing GOOG-CABR-00430662).

¹⁵⁸ *See* Apr. 21, 2022 Hrg. Tr. 204:4-6 (Ms. Sadowski: “UMA is very privacy preserving, which is why it’s the data set we use and we still track metrics even in Incognito mode.”); *id.* at 204:11-12 (“Q. Can UMA data be used to identify individual users? A. No. So UMA data is keyed by a Client ID, which roughly corresponds to a Chrome install. So you could have one user, like for example, myself, I have two different versions of Chrome running on my laptop, a Canary version, like an early version that might be buggy, and then the main, normal Chrome version. I also have a version of Chrome running on my phone, and all of those will have different ID’s associated with the same person, me. You could also have multiple people using the same Chrome install and have the same Client ID associated with multiple people.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

subset of users' UMA data), and it is not sufficiently reliable or accurate to use to identify class members.¹⁵⁹

- c. UMA data is designed for aggregate analysis and not for identifying specific users.¹⁶⁰
- d. UMA data is not joined with authenticated identifiers.¹⁶¹

¹⁵⁹ See GOOG-CABR-00057895, at -895 (“[E]xtrapolating [UMA metrics across all users] should only be used to give a ballpark estimate, as there are subtle biases in the UMA/UKM populations.”); GOOG-CABR-00057779, at -779 (“UMA data is only collected for users who have UMA enabled. This is controlled by a checkbox presented at installation, and through advanced Settings. After M55, UMA became opt-out on all platforms for new users, and Stable channel of Android and Windows Chrome further reduce the population, by randomly sampling a subset of clients.”).

¹⁶⁰ See Monsees Apr. 9, 2021 Tr. 212:10-16 (“[T]he user metrics analysis, my understanding is that for a signal like Chrome sync, which is a feature, right, within the -- the Chrome app itself, a [verbatim] engineer could configure UMA to say, ‘Pull me some aggregate statistics so I know how many Chrome device installs have sync enabled.’”); Svitkine Oct. 4, 2021 Tr. 159:15-17 (“UMA is not meant to compare like two specific Chrome instances. It is meant to look at aggregated data across many, many instances.”).

¹⁶¹ See Svitkine Tr. 44:20-22 (“[T]ak[ing] the existing UMA data and somehow join[ing] it with Gaia . . . there's many reasons why that can't be done.”); *id.* at 53:5-10 (“To be clear, the UMA system is not associated with a specific Gaia account with a specific user account in Chrome. It is a separate system that, as we previously discussed, cannot be joined with a user's e-mail address or their Gaia ID.”); Apr. 21, 2022 Hrg. Tr. 205:3-5 (“205:3-5 (Ms. Sadowski: “There isn't a way to directly join UMA data with Google account information. We do not have a mapping from UMA Client ID's to Google user names.”); *id.* at 205:23-25 (“Q. In your understanding as a Chrome engineer, can these bits found in these identified search logs be mapped to UMA data? A. No, they cannot.”); *id.* at 208:24-25 (“There's no key that you can join UMA with another data set that is a . . . User ID based data set.”); *see also* GOOG-CABR-00057918, at -918 (“[T]he primary metrics data sources for Chrome usage statistics (UMA and UKM) are not keyed by Gaia ID, and so are not joinable with other data about specific users (e.g. data from [REDACTED]). Furthermore, these two data sources are not joinable with each other. If we started keying UMA by Gaia ID, then the metrics would no longer be ‘anonymous’, which could mean that all previously acquired user consents for its collection would be invalid . . . This restriction means that anyone wanting to perform an analysis which joins, say, UMA data with some other data source is currently out of luck.”); GOOG-CABR-00056264, at -264 (“[T]he Chrome Data team is not planning on enabling gaia-keyed logging.” (emphasis in original)).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

131. Nor does the process of using activity and “timestamps” in conjunction with UMA data that Mr. Hochman references work to identify class members.¹⁶² Publicly available Chromium source code¹⁶³ describes the process by which timestamps are created:

```
// The timestamp for the event, in seconds.
// This value comes from Chromium's TimeTicks::Now(),
// which is an abstract
// time value that is guaranteed to always be
// non-decreasing (regardless of
// Daylight Saving Time or any other changes to the
// system clock).
// These numbers are only comparable within a session.
// To sequence events
// across sessions, order by the |session_id| from the
// ChromeUserMetricsExtension message.
```

132. Publicly available Chromium source code further explains that:¹⁶⁴

```
// `TimeTicks` and `ThreadTicks` represent an abstract
// time that is most of the
// time incrementing, for use in measuring time
// durations. Internally, they are
// represented in microseconds. They cannot be
// converted to a human-readable
// time, but are guaranteed not to decrease (unlike
// the `Time` class). Note
// that `TimeTicks` may "stand still" (e.g., if the
// computer is suspended), and
// `ThreadTicks` will "stand still" whenever the
// thread has been de-scheduled
// by the operating system.
```

¹⁶² Hochman ¶ 258.

¹⁶³ Chromium Code Search, “user_action_event.proto,”

“https://source.chromium.org/chromium/chromium/src/+main:third_party/metrics_proto/user_action_event.proto;l=21-27?q=time_sec (last visited June 3, 2022).

¹⁶⁴ Chromium Code Search, “time.h,”

<https://source.chromium.org/chromium/chromium/src/+main:base/time/time.h;l=16-22?q=TimeTicks> (last visited June 3, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

133. Mr. Hochman states that “[k]nowing that detailed event level data are stored with timestamps, IP address and user agent in these logs, it is my opinion that UMA can be joined with ads event logs.”¹⁶⁵

134. In my opinion, the timestamps Mr. Hochman references are not recorded in a way that can facilitate joining of UMA data with ads event logs. That is because the timestamp value “comes from Chromium’s TimeTicks,” which publicly-available Chromium source code shows are abstract values that cannot be compared across sessions. Thus, they are not, for example, a universal GMT time record that will apply across data sources, but are instead an abstract value that “cannot be converted to a human-readable time”¹⁶⁶ and not joined with universal time. Therefore, in my opinion, these records are not suitable for joining UMA and ads event logs.

b. PPID-Mapped Biscotti ID & Analytics User-ID

135. *First*, as discussed above (¶¶ 60-64), PPIDs and Analytics User-IDs are hashed first-party IDs generated, assigned, and used by some publisher websites. Google cannot readily use PPID-Mapped Biscotti IDs or Analytics User-IDs to identify class members for a number of reasons, including: (i) they are controlled by the publisher and are provided to Google by the publishers in an encrypted form that does not contain PII; (ii) not all publishers use PPIDs or Analytics User-IDs; (iii) different publishers may provide the same hashed value for the same user.

136. *Second*, based on my review of documents and data produced in this case, the PPID-mapped Biscottis and Analytics User-IDs are only keyed to a user’s activity on the

¹⁶⁵ Hochman ¶ 258.

¹⁶⁶ Chromium Code Search, “time.h,”

https://source.chromium.org/chromium/chromium/src/+/_main:base/time/time.h;l=16-22?q=TimeTicks (last visited June 3, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

specific website that has elected to use PPID or User-ID. Therefore, even if Mr. Hochman's method located the PPID-mapped Biscotti or Analytics User-ID of a purported class member, the only private browsing data associated with that ID would be from when that class member logged into the specific website that uses PPID or User-ID.

137. *Third*, Mr. Hochman also does not account for the prevalence of account-sharing in his claims regarding PPID-Mapped Biscotti IDs and Analytics User IDs. Sharing of accounts on third-party websites is a well-documented phenomenon that renders his methodology for identifying class members via PPIDs and Analytics User ID over-inclusive in certain common scenarios.¹⁶⁷ Consider, for example, a situation where User 1 and User 2 share a NYTimes.com account login and password in order to use features that require a user to log in to a NYTimes.com account (*e.g.*, to read articles that are protected by a paywall). Whenever User 1 or User 2 signs into NYTimes.com, the PPID will be the same for both users. If User 1 does not log in to NYTimes.com in Incognito mode, but User 2 does, then the shared PPID's presence in logs of both User 1 and User 2's activity will (incorrectly) point to User 1 as an Incognito mode user. Mr. Hochman does not account for this scenario, which is likely a common occurrence in light of the widespread prevalence of account-sharing. In my opinion, this failure would render Mr. Hochman's proposed use of PPIDs to identify class members as over-inclusive and inaccurate.

¹⁶⁷ See, *e.g.*, Alex Sherman, "Netflix estimates 100 million households are sharing passwords and suggests a global crackdown is coming," CNBC, <https://www.cnbc.com/2022/04/19/netflix-warns-password-sharing-crackdown-is-coming.html> (Apr. 20, 2022); Matt Richtel, "Young, in Love and Sharing Everything, Including a Password," N.Y. Times, <https://www.nytimes.com/2012/01/18/us/teenagers-sharing-passwords-as-show-of-affection.html> (Jan. 17, 2012) ("In a 2011 telephone survey, the Pew Internet and American Life Project found that 30 percent of teenagers who were regularly online had shared a password with a friend, boyfriend or girlfriend. The survey, of 770 teenagers aged 12 to 17, found that girls were almost twice as likely as boys to share. And in more than two dozen interviews, parents, students and counselors said that the practice had become widespread.").

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

c. Biscotti ID

138. Mr. Hochman proposes that “[i]f a user has signed into their Google Account in private browsing mode within the past 90 days, then Google can also retrieve the user’s Biscotti IDs from GAIA logs and use those Biscotti IDs to locate the user’s private browsing records while they are not signed into Google from the Biscotti logs.”¹⁶⁸

139. As an initial matter, a user who signs into a Google Account in private browsing mode is, by definition, outside of the scope of the proposed classes because the class is limited to users who “were not logged into their Google account.”¹⁶⁹ Therefore, Mr. Hochman is proposing identifying users who expressly fall outside of the class in the hope that they may also have some activity that meets the class definition.

140. But Mr. Hochman’s proposal will not work to identify those potential class members because the “Biscotti IDs from GAIA logs” that Mr. Hochman references will not be the same across separate private browsing mode sessions (discussed *supra* at [§ III.A.2](#)). Thus, the Biscotti IDs from a “signed-in Incognito session” cannot be used to “locate the user’s private browsing records” from a separate “signed-out Incognito session” because the Biscotti to which data from the two sessions is keyed will not be the same. Mr. Hochman does not explain how he would do so, and in my opinion, he cannot.

141. Additionally, Mr. Hochman’s proposed methodology does not account for device or account sharing (discussed *infra* at [§ III.J](#)) and suffers from the same defects identified above related to PPID and User ID (at [§ III.G.2](#)).

¹⁶⁸ Hochman ¶ 304.

¹⁶⁹ Complaint ¶ 192.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

H. Opinion 8: Mr. Hochman’s Opinion That The “maybe_chrome_incognito” Bit Reliably Detects Incognito Traffic (# 23) Is Incorrect

142. An important step in Mr. Hochman’s proposed method is to use the maybe_chrome_incognito bit to query Google’s logs to “find instances of users browsing within Incognito on [a] non-Google website while signed out of Google.”¹⁷⁰ As explained above, the industry guideline is for private browsing mode to not be detectable by websites. Google Chrome’s Incognito mode aligns with these guidelines, as Google has intentionally designed Chrome not to send an explicit signal to websites for the purpose of detecting Incognito mode.¹⁷¹

143. The *maybe_chrome_incognito* bit is a boolean field that relies on the absence of the X-Client-Data header to approximate and monitor traffic Google receives from Chrome instances in Incognito mode.¹⁷² In my opinion, the absence of the X-Client-Data header cannot

¹⁷⁰ Hochman ¶ 288.

¹⁷¹ See, e.g., GOOG-CABR-00059864, at -864 (“Today, some sites use an unintended loophole to detect when people are browsing in Incognito Mode. Chrome’s FileSystem API is disabled in Incognito Mode to avoid leaving traces of activity on someone’s device. Sites can check for the availability of the FileSystem API and, if they receive an error message, determine that a private session is occurring and give the user a different experience. With the release of Chrome 76 scheduled for July 30, the behavior of the FileSystem API will be modified to remedy this method of Incognito Mode detection. Chrome will likewise work to remedy any other current or future means of Incognito Mode detection.”); Adhya Tr. 188:18-21 (“Chrome does its best to make sure that -- and -- and it’s a principle around all the work they do around this space to make sure servers can’t detect Incognito Mode.”); Barb Palser, “Protecting private browsing in Chrome,” Google, <https://www.blog.google/outreach-initiatives/google-news-initiative/protecting-private-browsing-chrome/> (Jul. 18, 2019) (Chrome blogpost announcement disabling FileSystem API, which was used as a loophole to detect Incognito Mode (citing W3C, “W3C TAG Observations on Private Browsing Modes,” <https://w3ctag.github.io/private-browsing-modes/#features-supporting-private-browsing> (Apr. 9, 2020))).

¹⁷² GOOG-CABR-04470006, at -009 (“For Chrome, we have the x-client-data header in addition as a signal. But again, this will be heuristics-based, and can never [be] 100% accurate. The goal is to keep an eye on an envelop[e] of such traffic.”). Similarly, the “is_chrome_incognito” and “is_chrome_non_incognito” bits also rely on the absence of the X-Client-Data header. See Sadowski Tr. 76:5-16 (explaining that the value of the “is_chrome_non_incognito” bit is determined by whether an “X-Client-Data header [is] in the

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

be used to reliably detect Incognito traffic because there are a variety of cases in which the X-Client-Data header *is not* sent by a Chrome browser when a user is using a browser in non-Incognito mode (false positives).¹⁷³ Scenarios that can lead to false positives include the following:

- a. User has opted out of Chrome experiments.¹⁷⁴
- b. New browser instances.¹⁷⁵
- c. Browser not used for 30 days or more.¹⁷⁶
- d. Number of variation IDs in X-Client-Data header is too large.¹⁷⁷
- e. Population of X-Client-Data Header is blocked by a firewall.¹⁷⁸

144. Mr. Hochman contends that the false positives identified by Google witnesses “range from merely theoretical to exceedingly rare.”¹⁷⁹ Mr. Hochman does not explain how he arrived at this conclusion or make any attempt to quantify how often these scenarios occur.¹⁸⁰ As

request that is sent” and “that is all that it looks for.”); *id.* 77:20–78:2 (stating the “is_chrome_incognito” bit “is not derived through some other mechanism than presence or absence of X-Client-Data header, and suffers from the same limitations.”).

¹⁷³ See Berntson June 16, 2021 Tr. 374:12-376:11.

¹⁷⁴ GOOG-BRWN-00847297, at -298 (“However, Chrome doesn't send this header in Incognito mode (privacy concern) or the user opted-out in Chrome.”).

¹⁷⁵ Berntson June 16, 2021 Tr. 374:24-375:7 (“For the first case where there are instances where the X-Client-Data header is empty and it’s not in incognito browsing, there -- there are quite a few different ways that that can happen. One is if it’s a new browser instance, no X-client header is present in any call out from the browser.”).

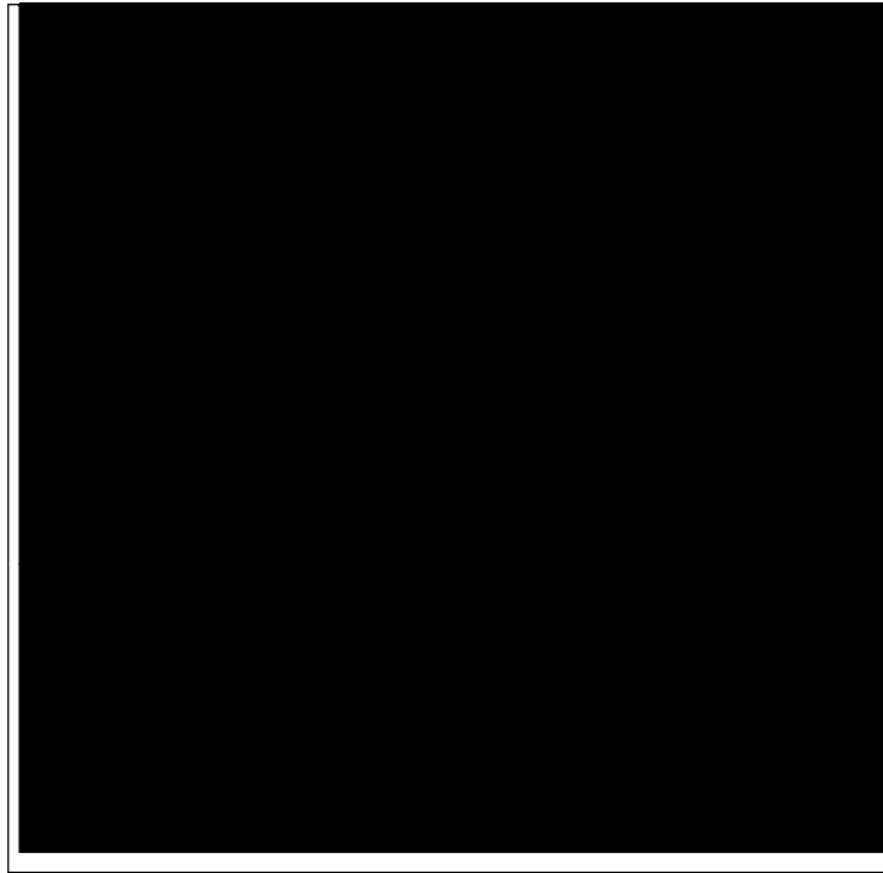
¹⁷⁶ *Id.* at 375:8-12 (“[I]f you haven’t used your browser for 30 days or more, the X-client header data is considered to be stale and just purged and no X-Client-Data header is passed.”).

¹⁷⁷ *Id.* at 375:13-19 (“Another case is if the variation IDs that are carried in the X-Client-Data header, if too many are returned to Chrome to prevent the requests coming from Chrome from being too large, they just delete them all and so you’d see no X-Client-Data header.”).

¹⁷⁸ *Id.* at 375:20-376:11 (“Yet another permutation is the presence of a firewall can also prevent Chrome, the browser, from getting the variation IDs that are used to populate the X-Client-Data header, and this is because the variation IDs are basically instructions as to what new features are enabled in the browser, and so Chrome, after it starts out, will make an asynchronous call to retrieve these data from the server, and if that server endpoint is blocked by a firewall, no X-Client-Data header is provided, none of the variation IDs. So that’s another case where you can have an empty X-Client-Data header.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

such, Mr. Hochman does not propose a method for accounting for false positives that would exclude false positives from the class. And as documents Google produced show, false positives based on analysis of UserAgent and the absence of the X-Client-Data header are not uncommon (and in some cases cannot be excluded):¹⁸¹



¹⁷⁹ Hochman Appendix G ¶ 27.

¹⁸⁰ As to the scenario where the X-Client-Data header is not sent because the number of variation IDs is too large, Hochman contends that Dr. Berntson testified that “when directly asked, ‘how often’ this occurs, he admitted that ‘I have not seen this flagged as a problem.’” Hochman Appendix G ¶ 30 (citing Berntson June 16, 2021 Tr. 385:4-5). However, Dr. Berntson did not testify that this does not occur—he merely noted that it is not a “problem” that Google has attempted to resolve. *See* Berntson June 16, 2021 Tr. 388:13-18 (“I am not aware of any bugs that are currently open that are describing this as a problem. Whether it’s actually a problem or not is different than whether or not it’s happening.”).

¹⁸¹ GOOG-CABR-04470006, at -016; *see also* Berntson June 16, 2021 Tr. 373:22-376:11; GOOG-BRWN-00845673, at -674 (“There is one condition in which the header also isn’t sent, which we cannot exclude: If the browser hasn’t received any experiment config yet, it will not send the header. This accounts for roughly 1.5-2% of browsing.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

145. Based on my experience in the industry and review of the documents and testimony provided by Google in this case,¹⁸² there is no way to identify and exclude the false positives in logs (and exclude logs associated with false positives from the logs used to identify the class) because the reason for false positives is not “observable from a server perspective.”¹⁸³ As such, it is my opinion that the `maybe_chrome_incognito` field can not be used to reliably detect Incognito traffic, let alone identify purported members of Class I.

I. Opinion 9: Mr. Hochman’s Proposal To Identify Class II (# 22) Is Unreasonable And Unreliable¹⁸⁴

146. Mr. Hochman’s proposal to identify class members for Class II (Non-Chrome Class)¹⁸⁵ consists of the following steps:

- a. Identifying the email addresses of all Google account holders.¹⁸⁶
- b. Sending an email notification “limited to email accounts associated with people in the United States, based on Google’s own records”¹⁸⁷ that, if

¹⁸² See, e.g., GOOG-CABR-04470006, at -009 (“For Chrome, we have the x-client-data header in addition as a signal. But again, this will be heuristics-based, and can never [be] 100% accurate. The goal is to keep an eye on an envelop[e] of such traffic.”); McClelland Tr. 166:11-19 (“Q. Wouldn’t the absence of the X-Client Data Header indicate a user is in Incognito mode? A. No, not necessarily. It’s not a strong enough signal to be confident that the user is in Incognito mode.”); Berntson June 16, 2021 Tr. 373:22-376:11.

¹⁸³ See Berntson June 16, 2021 Tr. 384:23-24.

¹⁸⁴ To the extent Mr. Hochman proposes using this methodology for identifying members of Class I (Chrome Class), I also disagree for the same reasons discussed in this section.

¹⁸⁵ Complaint ¶ 192 (“All non-Chrome browser users with a Google account who accessed a non-Google website containing Google tracking or advertising code using any such browser and who were (a) in “private browsing mode” on that browser, and (b) were not logged into their Google account on that browser, but whose communications, including identifying information and online browsing history, Google nevertheless intercepted, received, or collected from June 1, 2016 through the present (the “Class Period”).”).

¹⁸⁶ Hochman ¶ 296.

¹⁸⁷ *Id.* Mr. Hochman contends that “given the broad usage of the private browsing modes throughout the class period, it is [his] opinion that most of those people would be class members. If required, it would also be possible to use Google data to further limit email notification to accounts associated with some private browsing behavior.” *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

required before notification, would purportedly be limited to “accounts associated with some private browsing behavior” via Mr. Hochman’s proposed fingerprinting methods.¹⁸⁸

- c. After notification, users who seek to become class members could “provide additional information regarding their private browsing during the class period, or those users could provide certain identifying information to be used to search for records in Google’s data sources” that could purportedly “readily be used to verify whether that user was a Google Account holder during the class period and otherwise assess any response to any notifications.”¹⁸⁹ This “additional information” provided by users would include (i) an identification of non-Google websites they visited during the class period along with any available IP address and user agent information; (ii) if the IP address and user agent are not available to the user, this would be “obtained from Google’s own GAIA records” and then used to identify private browsing data via Mr. Hochman’s proposed fingerprinting methodology.¹⁹⁰
- d. “If a user has signed into their Google Account in private browsing mode within the past 90 days, then Google can also retrieve the user’s Biscotti IDs from GAIA logs and use those Biscotti IDs to locate the user’s private browsing records while they are not signed into Google from the Biscotti logs.”¹⁹¹

¹⁸⁸ *Id.*

¹⁸⁹ *Id.* ¶ 300.

¹⁹⁰ *Id.* ¶ 302.

¹⁹¹ *Id.* ¶ 304.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- e. Alternatively, tasking users with extracting cookies from the browser¹⁹² or non-Google sign-in IDs from third party websites¹⁹³ “with the help of technology professionals” while browsing in private mode and then using those cookies or identifiers to identify private browsing data in Google logs.
- f. As another alternative, configuring Google code on third party websites to provide a notification to users when they visit a page in private browsing mode.¹⁹⁴

147. In my opinion, this proposed method is unreasonable, unreliable, and will result in widespread mis-identification for the following reasons:

- ❖ **First**, sending an email notification to all Google account holders would be overly broad because, even if it were limited to account holders in the United States, that notification would be sent to many non-class members, and Mr. Hochman does not propose a methodology for limiting the notification to class members. *See infra* [§ III.I.1](#).
- ❖ **Second**, the steps that Mr. Hochman proposes for limiting Class II to private browsing mode users “after notification” would not reliably identify users of private browsing modes on non-Chrome browsers. *See infra* [§ III.I.2](#).

1. Mr. Hochman’s Proposed Email Notification To All Google Account Holders Is Overly Broad And He Does Not Propose A Workable Methodology For Limiting The Notification To Class Members

148. Mr. Hochman’s proposed email notification is overly broad, and he does not propose a method for limiting the scope of such a notification.

¹⁹² *Id.* ¶ 305.

¹⁹³ *Id.* ¶ 306.

¹⁹⁴ *Id.* ¶ 298.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

149. *First*, the proposed first step would be overly broad because a record of all Google account holders would include many accounts associated with users, entities, and other organizations that are not members of the class (where, for example, the user or entity associated with the account never used private browsing mode in the United States to visit a site using Google web-services). Sending an email notification to all Google account holders would be overly broad because only a minority of the recipients of such a notification would be eligible for membership in the class.

150. *Second*, Mr. Hochman's proposed second step attempts to address this overbreadth by limiting the list of Google account holders to "email accounts associated with people in the United States."¹⁹⁵ Mr. Hochman indicates that he believes this limitation would be sufficient because "given the broad usage of the private browsing modes throughout the class period, it is [his] opinion that most of those people would be class members."¹⁹⁶ As such, he appears to be assuming that "most" Google account holders who have used a non-Chrome browser have used that browser in private browsing mode. But he does not provide any support for this assumption, and user research shows that (i) most users *do not* use private browsing modes;¹⁹⁷ and (ii) at least for Chrome, a significant portion of users are not even aware of private browsing modes' existence.¹⁹⁸

151. *Third*, Mr. Hochman concedes that both of the proposed classes "are necessarily limited to individuals,"¹⁹⁹ and he states that the "notification could be limited to email accounts

¹⁹⁵ Hochman ¶ 296.

¹⁹⁶ *Id.*

¹⁹⁷ *See, e.g.*, GOOG-CABR-03751927, at -931 (only 35 percent of respondents indicated that they use private browsing mode).

¹⁹⁸ *See, e.g.*, GOOG-CABR-00422906, at -921 (20.3 percent of Chrome users surveyed reported that they never use Incognito mode and 11.7 percent were not even aware of Incognito mode at all).

¹⁹⁹ Hochman ¶ 295.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

associated with people in the United States.”²⁰⁰ In other words, he admits that the classes are composed of individuals, rather than, *e.g.*, organizations. But documents Google has produced show that Google accounts may be associated with organizations, rather than individual users.²⁰¹ Mr. Hochman does not propose a method for excluding these accounts from Class II (by, *e.g.*, explaining how the initial list of all Google account holders would be limited to “email accounts associated with people,” rather than email accounts associated with entities or organizations).

152. *Fourth*, Mr. Hochman states that “[i]f required, it would also be possible to use Google data to further limit email notification to accounts associated with some private browsing behavior,”²⁰² but he does not explain how he would limit the email notification in this way for Class II. In my opinion, there is no way to reliably identify users who used private browsing mode on non-Chrome browsers before notification because Google is unable to determine if browsing data received from non-Chrome browsers came from a user in private browsing mode,²⁰³ and Mr. Hochman’s proposed methodology does not explain how this could be done. To the extent he proposes that, *e.g.*, “individuals could provide additional information regarding their private browsing during the class period, or those users could provide certain identifying information to be used to search for records in Google’s data sources,” he states that this step would be performed “[a]fter notification.”²⁰⁴

²⁰⁰ Id. ¶ 295.

²⁰¹ GOOG-CABR-00086881, at -882 [REDACTED]

²⁰² Hochman ¶ 296.

²⁰³ See Berntson Mar. 18, 2022 Tr. 141:5-144:12; Monsees Apr. 9, 2021 Tr. 68:10-25; Monsees June 11, 2021 Tr. 452:20-454:19.

²⁰⁴ Hochman ¶ 300.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

2. Mr. Hochman's Proposed Methodology For Limiting Class II To Private Browsing Mode Users After Notification Is Unreliable

153. As to the steps that Mr. Hochman proposes for limiting Class II to private browsing mode users “after notification,” these methods are also unreliable and unreasonable. For example, Mr. Hochman proposes that “individuals could provide additional information regarding their private browsing during the class period, or those users could provide certain identifying information to be used to search for records in Google’s data sources.”²⁰⁵ As to the “additional information regarding their private browsing during the class period,” Mr. Hochman proposes that potential members of Class II would identify “any non-Google websites they visited during the class period” and provide “any available IP address and user agent information.” As discussed *supra* (at [§ III.G.1.](#)), attempting to identify records of signed-out private browsing via the use of IP address and user agent is not reliable because (i) the combination of IP address and user agent is not sufficient to identify an individual user (and in many cases will not even identify an individual device); and (ii) this proposed fingerprinting methodology would lead to false positives in light of widespread sharing of devices. Additionally, Mr. Hochman does not propose a methodology for users to confirm the non-Google websites that they visited in private browsing mode during the class period (if any), which cannot be determined by, *e.g.*, reviewing browsing history saved on a client because such histories are deleted when a private browsing session ends (by, for example, closing the only open private browsing window).

154. Mr. Hochman further proposes that “[i]f the IP address and user agent are not available, this information may be obtained from Google’s own GAIA records.”²⁰⁶ Google’s “GAIA records” show IP addresses and UAs associated with Google Account sign-ins, which

²⁰⁵ Hochman ¶ 300.

²⁰⁶ Hochman ¶ 302.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

could be, for example, an IP address and UA from a public library computer shared by hundreds of people. Therefore, this approach is not sufficiently reliable for the same reasons as Mr. Hochman's proposed method based on a user-supplied IP address and UA, discussed in [Section III.G.1](#). In addition, Mr. Hochman's proposed method does not account for the many different IP addresses and user agents that may be associated with a particular GAIA. Consider, for example, the 8,965 unique IP address and UA combinations associated with the seven GAIA IDs for the five named Plaintiffs from just the retained records that do not contain third-party confidential data.²⁰⁷ Mr. Hochman's proposed methodology would call for Google to search potentially billions of "orphaned" log entries for hundreds of different IP addresses and UAs for *each* Google account. And because the combination of IP address and UA is not sufficiently reliable to identify orphaned logs that correspond to the GAIA account holder, it would lead to over-inclusiveness and misidentification (where, *e.g.*, the IP address and UA from the GAIA records match the IP address and UA from logs of a *different user's* private browsing on a shared device).

155. I understand that pursuant to the Special Master process, Plaintiffs requested searches of 19 IPv4 + User Agent pairs and eight IPv6 + UserAgent pairs.²⁰⁸ The Special Master denied Plaintiffs' request for IPv4 + User Agent searches but allowed iterative searches using IPv6 + User Agent pairs, provided that Plaintiffs provide attestation that they are the only individuals who used the devices as indicated in their provided User Agent strings at the provided IPv6 addresses on a certain specific date and time.²⁰⁹ Plaintiffs then provided the

²⁰⁷ [Appendix F. IP Address + User Agent Data Analysis](#).

²⁰⁸ Consolidated 2022-04-08 3rd Round Search and Identifiers v2.

²⁰⁹ Apr. 14, 2022 Special Master Hrg. Tr. 83:22-84:24 ("IP addresses are not the named plaintiffs. It's the individuals themselves that we have to map to these actions. And if we can show that, that gives us the degree of confidence that allows us to move forward to, if there's a search result with that IPv6, with that user agent string, on that date and time, and we decrypt any of the IDs that are the result of that, it's improbable that it's anybody else other than the

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

required attestations for three new IPv6 + User Agent pairs they claim were copied from their devices.²¹⁰ Notably, none of the three IPv6 + User Agent pairs appears in the preserved data associated with seven GAIA IDs provided by the five named Plaintiffs that I reviewed, which contains 8,965 unique IP addresses + User Agent combinations.²¹¹ In my opinion, this indicates that Mr. Hochman's proposed IP + UA identification method is not viable.

156. Mr. Hochman also proposes using information from a user who "has signed into their Google Account in private browsing mode within the past 90 days" to "retrieve the user's Biscotti IDs from GAIA logs and use those Biscotti IDs to locate the user's private browsing records while they are not signed into Google from the Biscotti logs."²¹² This proposed method would not reliably identify members of Class II for the same reasons it would not reliably identify members of Class I, which is discussed *supra* [§ III.G.2.c](#). This proposal (i) would require decrypting Biscotti IDs that are always encrypted when stored with GAIA IDs (discussed *supra* [§ III.A.2](#)); and (ii) ignores the fact that each signed-out private browsing mode session will have a new Biscotti ID that will not match a Biscotti ID from a separate signed-out browsing session (and will not be stored in any GAIA logs if the user does not sign into a Google account). Because these signed-out private browsing mode Biscotti IDs will be distinct from any signed-in Biscotti IDs, it is not possible to "use those [signed-in] Biscotti IDs to locate the user's private browsing records while they are not signed into Google."²¹³

named plaintiff. ... So it's -- if we can narrow down on the IPv6s, because it's associated with the device, and show me confidence for each one of those sessions that are listed, you know, for the IP address UA -- what was that? Like about 27, 28? Yeah, that is, hey, Mr. Brown was the one doing it from that IPv6 address on this device, then we have some confidence that it's the person, not the machine.").

²¹⁰ 2022-05-09 - Plaintiff Brown Attestation; 2022-05-09 - Plaintiff Castillo Attestation; 2022-05-09 - Plaintiff Trujillo Attestation.

²¹¹ [Appendix F. IP Address + User Agent Data Analysis](#).

²¹² Hochman ¶ 304.

²¹³ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

157. Alternatively, Mr. Hochman proposes that “from any private browsing session (Incognito in Chrome or private browsing mode in other browsers), a user can (with the help of technology professionals) extract cookies from their browsers in private browsing mode (e.g., Biscotti cookies, GFP cookies and _ga cookies containing Analytics CIDs).”²¹⁴ According to Mr. Hochman, “[t]hese cookies are unique to that private browsing session and can be used to identify the user’s browsing activities in that session,” and, “[u]pon receiving these cookies, it is possible to locate the user’s private browsing data in Google logs and identify every website visited during that session that Google tracks.”²¹⁵

158. As explained above ([§ III.A.2](#)), unauthenticated identifiers set in a signed-out private browsing mode session are specific to a single session and are deleted once that session ends. In my opinion, this renders Mr. Hochman’s proposal unreasonable and unreliable for a number of reasons:

- a. *First*, Mr. Hochman’s proposed methodology could not be used to retrieve any identifiers from private browsing mode sessions that have already ended because those identifiers have already been deleted. Instead, he proposes that users could initiate a new private browsing session under the supervision of technology professionals and then use the results of this supervised session to establish membership in the class.
- b. *Second*, attempting to employ this methodology at scale (*i.e.*, for potentially millions of users) would not be feasible because it would require millions of individual users to each engage unspecified “technology professionals” and re-enact the testing that Mr. Hochman conducted for the named Plaintiffs.

²¹⁴ Hochman ¶ 305.

²¹⁵ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

159. As another alternative, Mr. Hochman proposes that “Google could require non-Google websites to provide pop-up notification whenever users visit websites for which Google will collect private browsing information.”²¹⁶ In my opinion, if Google were to require non-Google websites to provide such a pop-up notification, it would violate the W3C TAG guidelines discussed *supra* [§ III.B](#), and it would not identify prior private browsing activity.

160. As another alternative, Mr. Hochman proposes that “a user can extract their non-Google sign-in IDs (with the help of technology professionals) from websites that they visit in private browsing mode, including PPID, Analytics User ID and Google Ads User ID. Upon receiving these IDs, Google can locate the user’s private browsing data from those websites in logs.”²¹⁷

161. In my opinion, Mr. Hochman’s proposed use of signed-in identifiers on non-Google websites to identify class members is unreliable for the reasons discussed *supra* [§ III.G.2](#) and [§ III.A.2](#). Additionally, as discussed *infra* [§ III.G.2](#), Mr. Hochman’s proposed use of third party publisher sign-in IDs fails to account for shared accounts.

**J. Opinion 10: Mr. Hochman’s Proposed Methods For Identifying Class Members (# 22)
Do Not—And Cannot—Account For Shared Devices Or Accounts**

1. Mr. Hochman’s Proposed Methods Do Not Account For Shared Devices

162. As explained further below, Mr. Hochman’s proposed methods for identifying both classes do not account for shared devices, which renders his proposed methods over-inclusive for the following reasons:

- ❖ **First**, there is a significant body of research (including a published paper by Mr. Schneier, whom Plaintiffs have retained as an expert in this case) showing that

²¹⁶ Hochman ¶ 298.

²¹⁷ Hochman ¶ 306.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

sharing of devices is common. However, Mr. Hochman's methodology improperly assumes that identifying web browsing information from a device is sufficient to also identify the user of that device.

- ❖ **Second**, Incognito mode and other private browsing modes are specifically designed to provide on-device privacy for shared devices.
- ❖ **Third**, the failure to account for shared devices will lead to misidentification of class members based on the browsing activity of other users on these shared devices.

163. **First**, there is a significant body of research demonstrating that sharing of devices by more than one user is commonplace.²¹⁸ In fact, each of the named Plaintiffs testified

²¹⁸ See, e.g., Tara Matthews, et. al., “‘She’ll just grab any device that’s closer’: A Study of Everyday Device and Account Sharing in Households,” Proceedings of the ACM Conference on Human Factors in Computing Systems, ACM, <https://dl.acm.org/doi/pdf/10.1145/2858036.2858051> (2016), at 2 (“Among our key findings are that device and account sharing is common, and that mobile phones were shared as much as computers and more often than tablets.”); K. Levy and B. Schneier, “Privacy threats in intimate relationships,” Journal of Cybersecurity, <https://academic.oup.com/cybersecurity/article/6/1/tyaa006/5849222> (2020), at 10 (“[H]ouseholds are not units; devices are not personal; the purchaser of a product is not its only user.”); *id.* (criticizing the “assumption . . . that devices considered ‘personal’ are used by only one person” because “abundant research demonstrates that this is often not the case”); A. Brush and K. Inkpen, “Yours, Mine and Ours? Sharing and Use of Technology in Domestic Environments,” Proceedings of the 9th International Conference on Ubiquitous Computing, <https://www.microsoft.com/en-us/research/wp-content/uploads/2007/09/brushinkpenyoursmineours.pdf> (2007); B. Busse and M. Fuchs, “Prevalence of Cell Phone Sharing,” Survey Methods: Insights from the Field, <https://surveyinsights.org/?p=1019> (2013); H. Muller, J. Gove, and J. Webb, “Understanding Tablet Use: A Multi-Method Exploration,” Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services, <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38135.pdf> (2012).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

that they share their devices with others.²¹⁹ But Mr. Hochman does not propose a method for accounting for shared devices. Instead, his proposed methodology for identifying class members incorrectly assumes that there is a one-to-one ratio between devices and users.²²⁰

164. Publicly-available studies conducted by Google employees and other researchers confirm that users often share devices. For example, a 2016 research study found that “[m]any technologies rely on the assumption that they will be used by a single person,” but the “‘single-user’ assumption may not consider the many account and device sharing situations that arise during everyday life.”²²¹ Based on a study of U.S.-based households, the authors determined that “device and account sharing is common, and that mobile phones were shared as much as computers and more often than tablets—even though participants typically perceived phones as personal devices and did not realize how often they were shared until participating in our diary study.”²²²

²¹⁹ See, e.g., Brown Tr. 89:16-23 (Q. Does anyone else ever use any of your devices? A. From time to time . . . like my business partner might hop on my laptop or my brother or my girlfriend may hop on my phone.”); Trujillo Tr. 121:5-122-6 (“Okay. Do you share any of those devices you just mentioned with other people? A. Yes, I do, all three . . . It’s work computer, one -- one PC that’s used often, another PC that’s used not as often, and my mobile phone . . . Q. Okay. Has anyone ever used any of those devices to browse the internet? A. Yes.”); Byatt Tr. 239:25-240:4 (Q. Does anybody else have access to that laptop? A. Same as with the phone, where other people have certainly used it, but not -- not regularly.”); Davis Tr. 79:12-15 (“Q. And do -- do any other family members or friends ever use any of your devices? A. On very rare occasions. Like, if someone needs to get online to do something, I might let them.”); Castillo Tr. 260:17-22 (“Q. And your fiancé was upset at you because she learned of the -- A. I remember her being specifically annoyed that her surprise, her wedding engagement surprise was sitting on my computer and in her face every time she logged on to the computer.”).

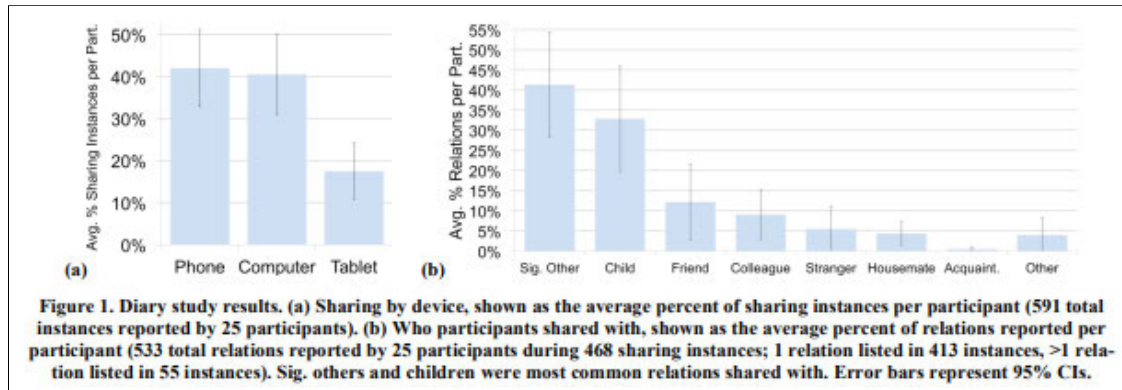
²²⁰ See, e.g., Hochman ¶ 293 (“[T]he IP addresses and User Agent strings can be used to join a user’s private browsing activities on non-Google websites with the user’s Google account identity, and Google could then notify the class member via the email address associated with that Google account.”).

²²¹ Tara Matthews, et. al., “‘She’ll just grab any device that’s closer’: A Study of Everyday Device and Account Sharing in Households,” Proceedings of the ACM Conference on Human Factors in Computing Systems, ACM, <https://dl.acm.org/doi/pdf/10.1145/2858036.2858051> (2016) (“Device Sharing Study”) at 5921.

²²² *Id.* at 5922.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

165. Their research yielded the following figures:²²³



166. The same study further determined that “‘personal’ devices are often shared; and sharing occurs in multiple ways and for a variety of reasons beyond the most obvious or visible.”²²⁴

167. Research surveys further confirm that device sharing is common. For example, in a 2015 study by the White House Council of Economic Advisers, “26 percent [of respondents] complain[ed] about having to share a computer with too many people in their household.”²²⁵

168. Other research on the prevalence of device sharing has reached similar conclusions, including a research paper authored by Plaintiffs’ expert Bruce Schneier.²²⁶ Mr. Schneier’s research notes that “People living in the same household may share computers,

²²³ *Id.* at 5925.

²²⁴ *Id.* at 5930.

²²⁵ Rick Paulus, “The Digital Divide Is About Much More Than Access,” *Pacific Standard*, <https://psmag.com/environment/digital-divide-more-complicated-than-access> (June 14, 2017).

²²⁶ See K. Levy and B. Schneier, “Privacy threats in intimate relationships,” *Journal of Cybersecurity*, <https://academic.oup.com/cybersecurity/article/6/1/tyaa006/5849222> (2020), at 1 (“[P]rivacy invasions by intimates are pervasive and deserving of focused study.”); *id.* at 2 (“In abusive partner situations, [technological privacy invasions by the abusive partner] can be a precursor to physical, emotional, and sexual abuse.”); *id.* (“People living in the same household may share computers, phones, and other connected devices.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

phones, and other connected devices.”²²⁷ His research further explains that the frequency with which devices are shared among users undercuts “[s]ystem designers[’] buil[t] in assumptions about intrafamilial privacy expectations,” which leads to incorrectly “treat[ing] a household as a ‘unit’ for purposes of information sharing.”²²⁸ Instead, Mr. Schneier urges that we should “realize that households are not units[,] devices are not personal[,] [and] the purchaser of a product is not its only user.”²²⁹ Based on my experience in the industry and review of the materials regarding device sharing cited above, I agree with Mr. Schneier’s conclusions on this point. However, Mr. Hochman does not make any attempt to account for these issues.

169. Indeed, Mr. Hochman’s proposed methodology for identifying class members equates “devices” with “users” in a way that assumes a one-to-one ratio between users and devices. This assumption is incorrect in light of the body of research determining that device sharing is common, including, *inter alia*, Mr. Schneier’s conclusion that “households are not units[,] devices are not personal[,] [and] the purchaser of a product is not its only user.”²³⁰ Further steps are required to identify specific *users* on shared devices. In my opinion, Mr. Hochman does not propose a method for doing so.

170. **Second**, documents Google produced in this case confirm that one of the primary reasons users use private browsing modes is to limit local storage of certain browsing information in order to prevent other users who may share their device from viewing locally-stored browsing histories.²³¹

²²⁷ *Id.* at 2.

²²⁸ *Id.* at 10.

²²⁹ *Id.*

²³⁰ *Id.*

²³¹ *See, e.g.*, GOOG-CABR-05171191, at -198 (noting that many users valued private browsing mode because [REDACTED]

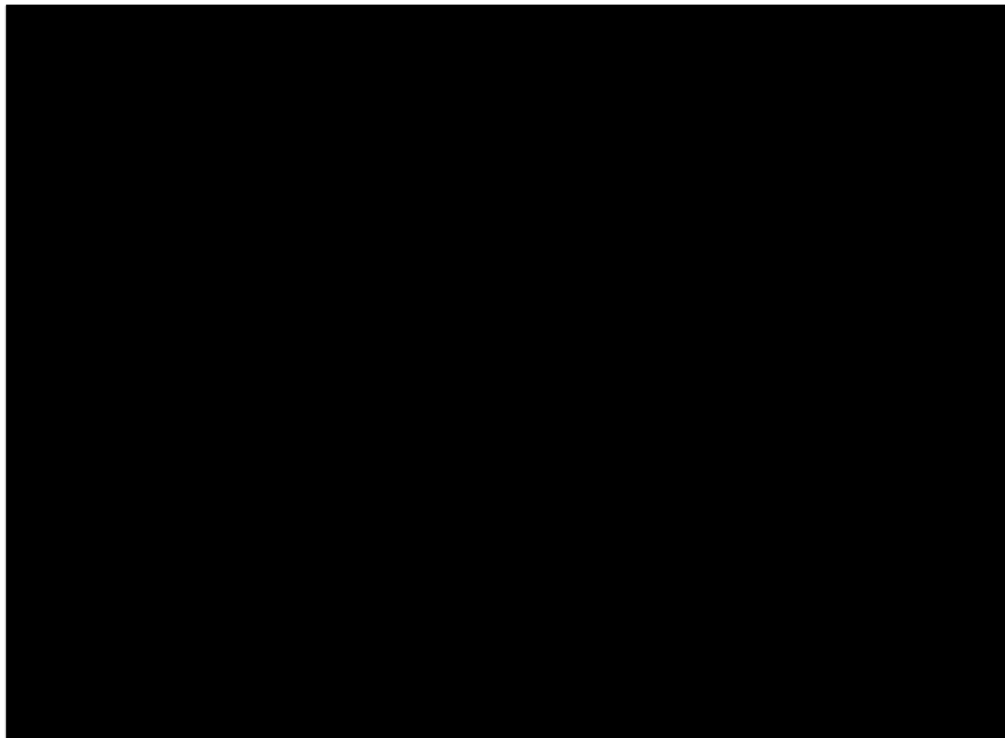
[REDACTED] id. at -212

[REDACTED] id.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

171. Indeed, the paradigmatic “engagement ring shopping” use case for Incognito mode is expressly directed to a shared device.²³² This involves a user using Incognito mode on a shared device to avoid his or her fiancé-to-be seeing browsing history related to engagement rings (or auto-filled URLs of websites where engagement rings are sold) when he or she uses the device. *See, e.g.*:

a. GOOG-BRWN-00166360, at -365:



²³² *See* GOOG-CABR-05171191, at -198 (noting that many users valued private browsing mode because

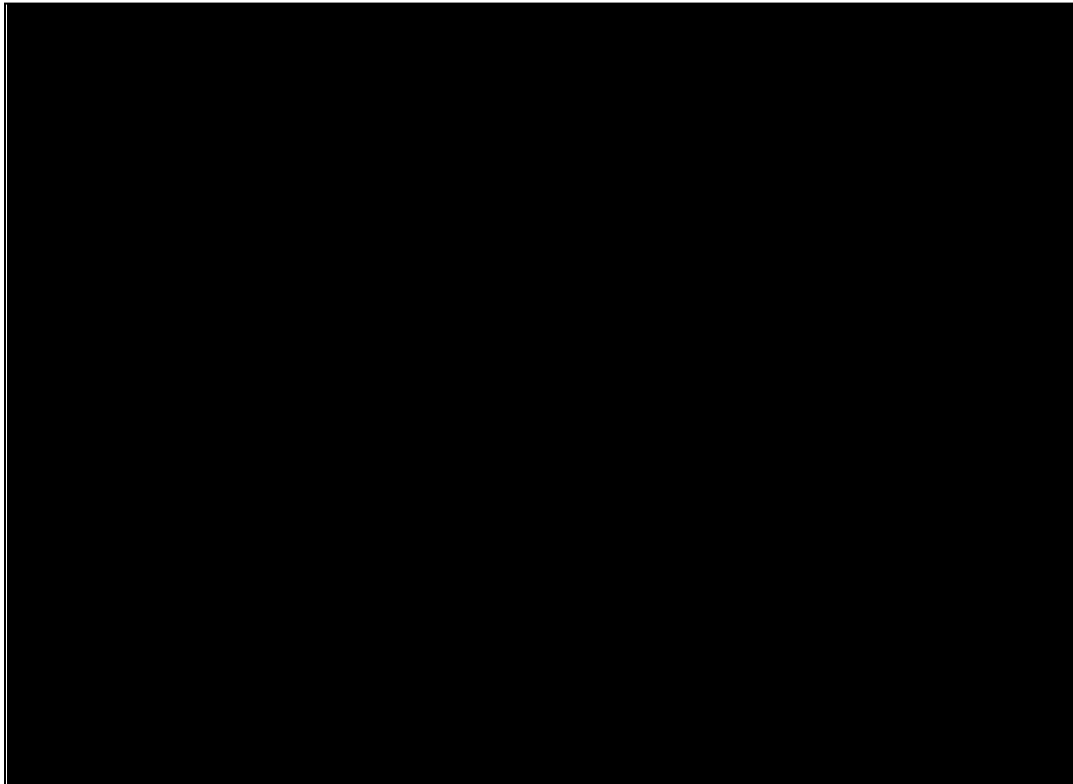
id. at -212

id.

see also GOOG-CABR-00148254 (user survey research report discussing variations between users’ reasons for using Incognito mode).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

b. GOOG-BRWN-00742713 at -887:



172. Google's internal initiatives also recognize that [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED].²³³ A number of social work organizations who provide resources for victims of domestic violence recommend the use of private browsing modes to limit an abuser's ability to view their partner's browsing history on a shared device.²³⁴

²³³ See GOOG-CABR-05468204.

²³⁴ See, e.g., Day One Services, "Maintaining Privacy When You Browse The Internet," <http://dayoneservices.org/be-safe/> (last visited June 3, 2022) (recommending use of private browsing to prevent local saving of browsing history); Sahara Cares, "Protect Yourself Online," <https://saharacares.org/protect-yourself-online/> (last visited June 3, 2022) ("If you are concerned about internet and browser safety, you have two options . . . Erase your browsing history after

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

173. **Third**, Mr. Hochman’s failure to account for shared devices will lead to misidentification of class members based on the browsing activity of other users on these shared devices. Mr. Hochman’s methodology for identifying members of Class I and Class II does not account for shared devices because it assumes that, by identifying a device, you have also conclusively identified the Google account holder who used that device in Incognito or other private browsing modes.²³⁵ Indeed, Mr. Hochman does not propose a methodology for identifying a *single user* based on an IP address and user agent that will be identical when multiple users share a single device. In my opinion, this necessarily makes Mr. Hochman’s proposed methodology over-inclusive and inaccurate.

174. Mr. Hochman’s proposed method for linking signed-out Incognito browsing activity to signed-in identifiers also does not account for shared devices.²³⁶ Consider, for

visiting this site [or] Use private browsing or incognito mode to browse this site so the visit is not logged in your history.”); WNY Postpartum Connection, Inc., “Domestic Abuse Support,” <https://www.wnypostpartum.com/domestic-violence-assistance> (last visited June 3, 2022) (“If you are in danger and believe someone could possibly be tracking your Internet usage, we highly recommend viewing this page in your browser’s incognito mode. You can do this by right-clicking the browser icon and selecting ‘New Incognito Window.’ If not using incognito, we recommend deleting your web history after viewing this page.”); Betty Griffin Center, “Safe Browsing,” <https://bettygriffincenter.org/help/> (last visited June 3, 2022) (“Your browser stores and records the webpages you visit on your computer/device. This is called a web history or a browser cache. You may want [to] conceal your activity online for safety reasons. There are a few ways to accomplish this . . . [for example,] Use private browsing. Most browsers have a privacy mode that allows you to visit websites without storing any record of your activity on your computer or device.”).

²³⁵ See, e.g., Hochman ¶ 291 (stating that “applying Google’s Incognito detection methodology using X-Client-Data and user agent accurately identifies Incognito data from the Second Iterative Search in the Special Masters process” without specifying how, in the case of a shared device, this data would be attributed to a particular user); ¶ 293 (stating that “the combination of IP address and User Agent” can be used to “attribute particular log entries to people” without addressing the reality that IP address and user agent will be identical for multiple “people” using the same browser on, e.g., a shared family computer).

²³⁶ See, e.g., Hochman ¶ 295 (“IP addresses and User Agent strings can be used to join a user’s private browsing activities on non-Google websites with the user’s Google account identity, and Google could then notify the class member via the email address associated with that Google account.”); ¶ 304 (“If a user has signed into their Google Account in private browsing mode within the past 90 days, then Google can also retrieve the user’s Biscotti IDs from GAIA logs

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

example, a desktop computer a family of four share. If (i) only one family member signs into a Google account on this shared device (the “Lone Signed In User”); and (ii) the IP address and user agent of the shared device is used to link private browsing information with the Lone Signed In User’s account, any use of Incognito mode by the other three family members would be attributed to the Lone Signed-In User and cause the Lone Signed-In User to be included in the class (even if he or she never personally used Incognito mode on the shared device). The proposed Chrome class is composed of *users* (not devices), but Mr. Hochman does not propose any way to determine which of the four family members in this example actually used Incognito mode on the shared family computer. In my opinion, the failure to account for shared devices makes Mr. Hochman’s proposed methodology for linking signed-out Incognito browsing activity to signed-in identifiers over-inclusive and inaccurate.

175. As another example, consider a publicly-available computer available for visitors to use in, *e.g.*, a public library. This shared device does not require users to log in to use the computer (or log out when they are finished). Over the course of a day, one hundred users use this machine at various points. If, for example, only one user uses Incognito mode (or another private browsing mode) on the same day, Mr. Hochman does not propose any way to determine which of these one hundred users actually used Incognito mode on this publicly-available device (where all one hundred users will have the exact same user agent and IP address). Indeed, if (i) a user logged into his or her Google account (the “Logged In User”) while using the device (and did not use Incognito mode at all), and (ii) the next user (the “Subsequent Incognito Mode User”) did not log into his or her own Google account, but proceeded to open an Incognito mode window and browsed a non-Google website that uses Google Analytics or

and use those Biscotti IDs to locate the user’s private browsing records while they are not signed into Google from the Biscotti logs.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Ad Manager code, then (iii) the proposed methodology would lead to a false positive because it would incorrectly identify the Logged In User as a class member based on the Subsequent Incognito Mode User's browsing activity on the shared machine, as the Logged In User's activity keyed to his GAIA ID would contain the same IP address and user agent as the Subsequent Incognito Mode User's browsing activity keyed to an "orphaned" Biscotti ID. Mr. Hochman's proposed methodology does not account for this scenario, which is not uncommon.²³⁷

176. Mr. Hochman's methodology for identifying class members by sending notifications to Google account holders, whose accounts are associated with some private browsing activity, also fails to account for shared devices because it relies on browsing activity on the same device (without providing any way to confirm that the browsing activity is actually attributable to the same Google account holder, rather than a different individual who used the device). Consider, for example, a laptop used by two different members of the same household. One of the users of this shared device (call him User_1) signed into a Google account on the shared machine once but never enabled private browsing. Another user of the shared device (User_2) uses the device exclusively for private browsing and has never logged in to a Google account on that shared device. Mr. Hochman's proposed methodology would result in a notification being sent to User_1, not the actual class member who browsed in private mode, User_2. In the domestic violence use case of private browsing mode mentioned above, the abuser may get notification based on the private browsing activities of the victim.

²³⁷ See, e.g., Google Help Page: Sign out of Chrome, <https://support.google.com/chrome/answer/9159867?hl=en&co=GENIE.Platform%3DDesktop> (last visited June 7, 2022) ("Sign out remotely . . . You can remove your account from one of your devices, even if you don't have that device with you. You'll be signed out from any computer you've used before, including the one you're using now.").

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

177. By failing to account for shared devices, Mr. Hochman's proposed methods for identifying members of both the Chrome Class and the non-Chrome Class are over-inclusive because they identify individual users as class members based on other users' browsing activity on the same device (where their user agent and IP address will be the same). In my opinion, he does not propose a methodology for excluding these cases, which, as discussed above, are quite common.

2. Mr. Hochman's Proposed Methods Do Not Account For Shared Accounts

178. Mr. Hochman's proposed methodology for identifying class members does not account for shared accounts, which is also common.²³⁸ Mr. Hochman's proposed methodology for using "Biscotti encryption keys" to "locate additional user's [sic] signed-out private browsing activities via their GAIA IDs for users who may have, on occasion, signed into their Google account in a private browsing session"²³⁹ (i) does not account for the prevalence of account sharing; and (ii) fails to propose a method for determining which user (or users) of a shared account would be a class member in situations where a subset of users with access to the shared account did not use private browsing modes during the class period. For example, consider an account shared between two grandparents with limited technical ability and a

²³⁸ See, e.g., K. Levy and B. Schneier, "Privacy threats in intimate relationships," *Journal of Cybersecurity*, <https://academic.oup.com/cybersecurity/article/6/1/tyaa006/5849222> (2020), at 2 ("Intimates might share social media and email accounts—and even if they have separate accounts, they may know one another's passwords."); C. Park, et al., "Share and share alike? An exploration of secure behaviors in romantic relationships," *Fourteenth Symposium on Usable Privacy and Security (SOUPS)*, https://www.researchgate.net/publication/325608530_Share_and_Share_Alike_An_Exploration_of_Secure_Behaviors_in_Romantic_Relationships (2018).

²³⁹ Hochman ¶ 244.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

grandchild in their care who is more technically capable.²⁴⁰ The Google account in this scenario is registered only to the grandfather, who has never used Incognito mode (and is in fact not aware of the existence of this feature at all).²⁴¹ If the grandchild used Incognito mode, and at least once logged into the grandfather's Google account in Incognito mode during the class period, but the grandparents did not, Mr. Hochman's proposed methodology would nonetheless include the grandfather in the class by joining the grandchild's private browsing mode activity with the grandfather's Google account identifier.

179. Similarly, Mr. Hochman's proposed methodology for identifying class members via PPIDs fails to account for sharing of accounts on third party websites. Third party websites that use PPIDs may require or allow users to log into their site for a number of reasons, including for example, to access "premium" content which is only made available to paid subscribers (*e.g.*, paywall-protected articles on NYTimes.com or washingtonpost.com). Sharing of login and password information by more than one user to access premium content is common.²⁴² Like shared devices, Mr. Hochman does not attempt to account for this practice.

180. Mr. Hochman's methodology also fails to account for GAIA IDs that are not associated with a specific individual. Documents Google produced indicate that [REDACTED]

²⁴⁰ See, *e.g.*, Juliette Garside, "Ofcom: six-year-olds understand digital technology better than adults," *The Guardian*, <https://www.theguardian.com/technology/2014/aug/07/ofcom-children-digital-technology-better-than-adults> (Aug. 6, 2014) ("[T]he average six-year-old child understands more about digital technology than a 45-year-old adult."); Laveh Waddel, "Will Today's Kids Be Stumped by the Technology of the Future?," *The Atlantic*, <https://www.theatlantic.com/technology/archive/2016/01/will-todays-kids-be-stumped-by-the-technology-of-the-future/425082/> (Jan. 26, 2016).

²⁴¹ See GOOG-CABR-00422906, at -068 (20.3 percent of survey respondents reported that they never use Incognito mode, and 11.7 percent were not aware of the feature).

²⁴² See, *e.g.*, Alex Sherman, "Netflix estimates 100 million households are sharing passwords and suggests a global crackdown is coming," *CNBC*, <https://www.cnbc.com/2022/04/19/netflix-warns-password-sharing-crackdown-is-coming.html> (Apr. 19, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

[REDACTED]

[REDACTED]

[REDACTED]²⁴³ Mr. Hochman does not propose a method for, *e.g.*, (i) identifying these multi-user accounts; and (ii) limiting the inclusion of class members to individual users of such accounts who actually used private browsing modes. Thus, if a single user of such an account used private browsing mode on a device with the same IP address and user agent as the multi-user account, applying Mr. Hochman’s methodology would include “entire businesses, groups of users and even services” in the class. This failure to account for multi-user accounts thereby renders Mr. Hochman’s proposed methodology over-inclusive.

²⁴³ GOOG-CABR-00086881, at -882.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

IV. REBUTTAL TO MR. SCHNEIER**A. Opinion 11: Mr. Schneier's Assertion That "Browsing Information Is Unique For Each User" Is Unsupported And Misleading**

181. According to Mr. Schneier: "Browsing Information Is Unique For Each User."²⁴⁴

182. In support of this claim, Mr. Schneier states that:

- a. "American citizens are justified in taking measures to minimize access to their browsing information, since it can be used to identify them. A 2013 study of 368,284 Internet users detected a unique browsing history for 69% of participants, and found that out of users for whom at least four visited websites were detected, 97% could be uniquely identified by their browsing history."²⁴⁵
- b. "This browsing information is a rich target for those online businesses that deploy CSSbased detection techniques to collect it. (CSS is an initialism for cascading style sheets, which are used to format web pages.) An attacker can ascertain URLs visited by a target's browser through applying CSS styles that differentiate visited and unvisited links. A study of results obtained from over a quarter-million web users found that over 94% of Google Chrome users were vulnerable to CSS-based browser history detection by sites they visited; a test of popular websites detected an average of 62.6% visited locations per client."²⁴⁶
- c. "A 2015 research paper illustrated how third-party cookies can be used by eavesdroppers— these are people who are not the owners of the websites visited or the cookies issued and used— to track people on the Internet. Simulating users browsing the web, the authors found that 'the adversary can reconstruct 62–73% of a typical user's browsing history.'"²⁴⁷

²⁴⁴ Schneier at 27.

²⁴⁵ Schneier ¶ 97 (citing L. Olejnik, C. Castelluccia and A. Janc, "Why Johnny can't browse in peace: On the uniqueness of web browsing history patterns," *Annals of Telecommunications* 1-2, <https://hal.inria.fr/file/index/docid/747841/filename/johnny2hotpet-finalcam.pdf> (June 2013)).

²⁴⁶ Schneier ¶ 98 (citing A. Janc and L. Olejnik, "Web browser history detection as a real-world privacy threat," *ESORICS'10: Proceedings of the 15th European Conference on Research in Computer Security*, <http://cds.cern.ch/record/1293097/files/LHCb-PROC-2010-036.pdf> (September 20, 2010)).

²⁴⁷ Schneier ¶ 99 (citing S. Englehardt, et al., "Cookies that give you away: The surveillance implications of web tracking," *WWW '15: Proceedings of the 24th International Conference on World Wide Web*, https://senglehardt.com/papers/www15_cookie_surveil.pdf (May 18, 2015)).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

183. As explained further below, Mr. Schneier’s assertion is unsupported and particularly misleading for the Data At Issue for the following reasons:

- ❖ **First**, the study cited by Mr. Schneier (§ 97) concluded that only 50 percent of the “Google profiles” that the authors analyzed when they limited their analysis to “sites on which [they] detected scripts from Google” were unique.²⁴⁸ As such, the remaining 50 percent of “Google profiles” were *not unique*. This research also did not conclude that 100 percent of any other browsing information was unique. Thus, Mr. Schneier’s assertion that “Browsing Information Is Unique For Each User” (i) is not supported by this study, which did not conclude that browsing information is in fact unique for each user because it did not determine that any set of browsing information or profiles were 100 percent unique; and (ii) is particularly misleading for the Data At Issue because the authors of that study used different data (*i.e.*, URLs provided to the researchers via an experimental system), and they determined that only half of the “profiles” they generated—using a different methodology and a different definition of “profiles” than any proposed by Plaintiffs—from visits to sites with embedded Google scripts were “unique.”
- ❖ **Second**, the security study cited by Mr. Schneier regarding potential vulnerabilities to “CSS-based browser history detection” (§ 98) merely identifies a potential vulnerability and does not address whether or not particular browsing histories are “unique,” let alone show that each user has a unique browsing history.²⁴⁹

²⁴⁸ Lukasz Olejnik, Claude Castelluccia and Artur Janc, “Why Johnny can’t browse in peace: On the uniqueness of web browsing history patterns,” *Annals of Telecommunications* at 13 <https://hal.inria.fr/file/index/docid/747841/filename/johnny2hotpet-finalcam.pdf> (June 2013).

²⁴⁹ S. Englehardt, et al., “Cookies that give you away: The surveillance implications of web tracking,” *WWW ‘15: Proceedings of the 24th International Conference on World Wide Web*, https://senglehardt.com/papers/www15_cookie_surveil.pdf (May 18, 2015).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- ❖ *Third*, the research paper cited by Mr. Schneier (¶ 99) regarding third-party cookie “eavesdroppers” addresses potential vulnerabilities and modeling of potential reconstruction of browsing histories, rather than analyzing the uniqueness of actual histories.²⁵⁰

184. In my opinion, the study cited by Mr. Schneier regarding the relative uniqueness of browsing histories does not support his claim that browsing information is “unique for each user” because the authors did not find that 100% of the browsing profiles analyzed were unique (*i.e.*, the “only one” or “able to be distinguished from all others of its class or type”).²⁵¹ And Mr. Schneier’s characterization of this study as supporting his conclusion is particularly misleading for the Data At Issue because (i) the authors of that study used different data collected via a different methodology (*i.e.*, URLs provided to the researchers via an experimental system); and (ii) they determined that only half of the “profiles” they generated—using a different methodology and a different definition of “profiles” than any proposed by Plaintiffs—from visits to sites with embedded Google scripts were “unique.” Mr. Schneier does not account for these methodological differences or address the authors’ conclusion that only 50 percent of the “profile” histories from sites using Google scripts that they generated were unique.²⁵²

²⁵⁰ *Id.*

²⁵¹ “Unique” means “the only one” or “able to be distinguished from all others of its class or type.” Merriam-Websters.com Dictionary, “Unique,” <https://www.merriam-webster.com/dictionary/unique#:~:text=1%20%3A%20being%20the%20only%20one%20of%20its%20kind%20Every%20snowflake,from%20Merriam%2DWebster%20on%20unique> (last visited June 6, 2022).

²⁵² L. Olejnik, C. Castelluccia and A. Janc, “Why Johnny can’t browse in peace: On the uniqueness of web browsing history patterns,” *Annals of Telecommunications* at 13 <https://hal.inria.fr/file/index/docid/747841/filename/johnny2hotpet-finalcam.pdf> (June 2013)).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

185. It is also my opinion that the study Mr. Schneier cites regarding CSS vulnerabilities²⁵³ does not support his conclusion that browsing information is unique for each individual user because the authors simply analyzed the susceptibility of certain users of internet browsers to attacks designed to obtain their browsing histories, and they did not evaluate the relative uniqueness of the histories. Whether or not a vulnerability could be exploited to obtain browsing information is orthogonal to a determination of whether or not the information that may be obtained by an attacker is unique for each user.

186. It is also my opinion that the study Mr. Schneier cites regarding potential “eavesdropping” via accessing third party cookies²⁵⁴ does not support his conclusion that browsing information is unique for each user because this research merely describes certain techniques that might be used by bad actors to obtain cookie values and unencrypted PII. Whether or not a certain attack might be suitable for a bad actor to obtain information is a distinct question from a determination of whether or not browsing information is unique for each user.

B. Opinion 12: Mr. Schneier’s Claim That “Personal Data Is Difficult To Anonymize And Easy To De-Anonymize” Is Unsupported And Is Incorrect For The Data At Issue

187. According to Mr. Schneier: “Personal Data Is Difficult to Anonymize and Easy to De-anonymize.”²⁵⁵ In support, Mr. Schneier:

- a. Claims that maintaining online anonymity is purportedly difficult “in the face of a focused and determined investigation.” In support of this claim, Mr. Schneier states that “[e]ven a team of highly trained Israeli assassins

²⁵³ A. Janc and L. Olejnik, “Web browser history detection as a real-world privacy threat,” ESORICS’10: Proceedings of the 15th European Conference on Research in Computer Security, <http://cds.cern.ch/record/1293097/files/LHCb-PROC-2010-036.pdf> (September 20, 2010)).

²⁵⁴ S. Englehardt, et al., “Cookies that give you away: The surveillance implications of web tracking,” WWW ‘15: Proceedings of the 24th International Conference on World Wide Web, https://senglehardt.com/papers/www15_cookie_surveil.pdf (May 18, 2015).

²⁵⁵ Schneier at 41.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

was quickly identified in Dubai, based on surveillance camera footage from around the city.”²⁵⁶

- b. Asserts that “[m]ost techniques for anonymizing data don’t work, and ostensibly anonymized data can be de-anonymized with surprisingly little information.”²⁵⁷ In support of this claim, Mr. Schneier cites the following:
 1. A 1997 study and follow up research from 2000 and 2006 in which the researcher deanonymized certain records by combining zip codes with gender and birth date records from census data or other data (e.g., healthcare records).²⁵⁸
 2. A 2006 study analyzing the effectiveness of attempted anonymization of “three months of search data for 657,000 users: 20 million searches in all” released by American Online (“AOL”).²⁵⁹
 3. A 2008 study of “10 million movie rankings by 500,000 anonymized customers” published by Netflix from which researchers were able to determine the identity of certain

²⁵⁶ Schneier ¶ 143 (citing Natasha Lomas, “France fines Google \$120M and Amazon 42M for dropping tracking cookies without consent,” Tech Crunch, <https://techcrunch.com/2020/12/10/france-fines-google-120m-and-amazon-42m-for-droppingtracking-cookies-without-consent> (Dec. 10, 2020)).

²⁵⁷ Schneier ¶ 144 (citing Paul Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization,” UCLA Law Review 57, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 (August 13, 2009)).

²⁵⁸ Schneier ¶ 145 (citing Latanya Sweeney, “Simple demographics often identify people uniquely,” Carnegie Mellon University Data Privacy Working Paper 3, <https://dataprivacylab.org/projects/identifiability/paper1.pdf> (2000); Philippe Golle, “Revisiting the uniqueness of simple demographics in the U.S. population,” 5th ACM Workshop on Privacy in the Electronic Society (WPES’06), Alexandria, Virginia, <https://crypto.stanford.edu/~pgolle/papers/census.pdf> (Oct. 30, 2006); L. Sweeney, A. Abu and J. Winn, “Identifying participants in the Personal Genome Project by name (A re-identification experiment),” arxiv.org, <https://arxiv.org/abs/1304.7605> (2013); Latanya Sweeney, “Only you, your doctor, and many others may know,” Technology Science 2018, <https://techscience.org/a/2015092903> (September 28, 2015); Ji Su Yoo, et al., “Risks to patient privacy: A re-identification of patients in Maine and Vermont statewide hospital data,” Technology Science 2018, <https://techscience.org/a/2018100901> (Oct. 8, 2018); Katherine E. Boronow, et al., “Privacy risks of sharing data from environmental health studies,” Environmental Health Perspectives 128, no. 1, <https://ehp.niehs.nih.gov/doi/10.1289/EHP4817> (Jan. 2020)).

²⁵⁹ Schneier ¶ 146 (citing M. Barbaro and T. Zeller Jr., “A face is exposed for AOL Search No. 4417749,” New York Times, <http://www.nytimes.com/2006/08/09/technology/09aol.html> (Aug. 9, 2006)).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

individuals “by comparing rankings and time stamps with public rankings and time stamps in the Internet Movie Database.”²⁶⁰

4. A 2015 study analyzing the relative anonymity of certain credit card data.²⁶¹
 5. A 2019 study analyzing certain demographic data.²⁶²
 6. A 2018 study analyzing the volume of data sent to Google’s servers in connection with the use of certain Google web applications (e.g., Gmail) and browsing on websites that use Google Analytics.²⁶³
 7. Studies analyzing attempts to deanonymize data sets by combining them with location data or school records.²⁶⁴
 8. Mr. Schneier’s assertion that (i) telephone records could be partially deanonymized by correlating them with a database of telephone order information; and (ii) “Amazon’s online book reviews could be the key to partially de-anonymizing a database of credit card purchase details.”²⁶⁵
- c. Claims that “Joinability is a risk whether or not data is actually being joined,” and Google could purportedly use its “database of users’ Internet searches . . . [to] de-anonymize a public database of Internet purchases, or

²⁶⁰ Schneier ¶ 147 (citing A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” 2008 IEEE Symposium on Security and Privacy, Oakland, California, <https://web.stanford.edu/class/cs245/win2020/readings/netflix-deanonymization.pdf> (May 18-20, 2008)).

²⁶¹ Schneier ¶ 148 (citing Yves-Alexandre de Montjoye, et al., “Unique in the shopping mall: On the re-identifiability of credit card metadata,” *Science* 347, no. 6221, <https://www.science.org/doi/full/10.1126/science.1256297> (Jan. 30, 2015)).

²⁶² Schneier ¶ 149 (citing L. Rocher, J. Jendrickx and Y. de Montjoye, “Estimating the success of re-identifications in incomplete datasets using generative models,” *Nature Communications* 10, <https://www.nature.com/articles/s41467-019-10933-3> (July 23, 2019)).

²⁶³ Schneier ¶ 150 (citing Douglas C. Schmidt, et al., “Google data collection,” Vanderbilt University, <https://digitalcontentnext.org/wpcontent/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf> (Aug. 15, 2018)).

²⁶⁴ Schneier ¶ 151 (citing Dániel Kondor, et al., “Towards matching user mobility traces in large-scale datasets,” arXiv:1709.05772, <https://arxiv.org/pdf/1709.05772.pdf> (Aug. 13, 2018); Eli Jacobson, et al., “De-identification is insufficient to protect student privacy, or What can a field trip reveal?” *Journal of Learning Analytics* 8, no. 2, <https://www.learning-analytics.info/index.php/JLA/article/view/7353> (2021)).

²⁶⁵ Schneier ¶ 152.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

zero in on searches of medical terms to de-anonymize a public health database.”²⁶⁶

- d. Characterizes certain documents produced by Google as purportedly containing admissions that the Data At Issue can be joined to signed-in authenticated Google data that support Mr. Schneier’s opinion regarding joinability.²⁶⁷

188. As explained further below, Mr. Schneier’s assertion is unsupported by the sources he cites and incorrect for the Data At Issue for the following reasons:

- ❖ **First**, Mr. Schneier cite examples of purported difficulties associated with anonymization that do not address the Data At Issue, which is readily distinguishable from, *e.g.*, surveillance camera footage of Israeli assassins, birth dates, ZIP codes, voter registration databases, AOL Search data, Netflix movie rankings, publicly-posted IMDB ratings, credit card numbers, “user location data . . . combined with anonymized credit card data,” “telephone records . . . correlat[ed] . . . with a catalog merchant’s telephone order database,” Amazon online book reviews, and “Google[’s] database of users’ Internet searches.”
- ❖ **Second**, the Google documents that Mr. Schneier characterizes as “admissions” do not support his contentions because, *inter alia*, (i) they involve discussions of theoretical possibilities in a hypothetical world without Google’s policies and server-side architecture that prevent deanonymization of the Data At Issue; and (ii) these statements do not support his conclusion that “Google can connect individuals to private browsing sessions.”

²⁶⁶ Schneier ¶ 153 (citing Pern Hui Chia, et al., “KHyperLogLog: Estimating reidentifiability and joinability of large data at scale,” Proceedings of the IEEE Symposium on Security and Privacy, <https://milinda-perera.com/pdf/CDPSLDWG19.pdf> (2019)).

²⁶⁷ Schneier ¶ 155 (citing GOOG-BRWN-00705010; GOOG-CABR-05270014; Mardini Nov. 24, 2021 Tr. 346-347).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- ❖ **Third**, Google documents and testimony produced in this case show that Mr. Schneier’s assertion that “data is difficult to anonymize and easy to de-anonymize” is incorrect for the Data At Issue in light of Google’s successful efforts to anonymize the Data At Issue and prevent its re-identification.

189. In my opinion, the video surveillance story that Mr. Schneier cites does not support any claim that the Data At Issue is “difficult to anonymize” and “easy to deanonymize” because that is a wholly different type of data and analysis than the instant case (*i.e.*, surveillance camera footage of a few Israeli agents analyzed to identify a match for facial images in a database, rather than unauthenticated private browsing data that Plaintiffs claim can be reliably tied to the identity of millions of individual users). The Data At Issue does not include surveillance camera footage, nor does analysis of such data involve attempts to match images of individuals with such footage.

190. It is also my opinion that the vast majority of the studies that Mr. Schneier cites do not support his opinion in this case because they address categories of information that are different from the Data At Issue. For example, the Data At Issue does not include (i) birth dates, zip codes, or medical records; (ii) search requests that users input into Google’s search engine (or other search engines that do not use Google services like Bing or DuckDuckGo); (iii) Netflix movie rankings; (iv) credit card numbers; (v) demographic data (such as users’ age, gender and marital status); (vi) school records tied to location data; (vii) telephone records; or (viii) Amazon review and purchase histories. Importantly, the Data At Issue is also not merely anonymized by, *e.g.*, obscuring the identity of the individual user with whom the Data At Issue is associated. By contrast, the Data At Issue is never “authenticated” or otherwise tied to a specific user’s identity in the first place. *See supra* [§ III.A.1](#). And as discussed above, it cannot

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

be reliably tied to a specific user via Plaintiffs' proposed fingerprinting methodologies. *See supra* §§ III.G.1; III.F.1.

191. As to the lone study regarding certain Google services that Mr. Schneier cites,²⁶⁸ it is also my opinion that this research does not support Mr. Schneier's claim that the Data At Issue is difficult to anonymize because the researchers tested the exact opposite scenario that is at issue here: the researchers tested the potential for Google to associate data from a private browsing mode session with a user's identity *when a user signs into a Google account or service*.²⁶⁹ In this situation, the authors determined that Google can use a signed-in identifier received *when a user logs into Google* to associate signed-out activity with the user's Google account. For the Data At Issue, no such signed-in identifier will be sent to Google because Plaintiffs allege that the putative class members have *not signed into a Google account*, and thus the results of this test do not apply.

192. Additionally, the Google-produced documents that Mr. Schneier characterizes as "admissions" do not support his claims regarding joinability for the following reasons:

- a. GOOG-BRWN-00705010: Mr. Schneier contends that this document shows that "Google employees have admitted that 'it is possible for Google to join regular and Incognito sessions,'" and Google "can [therefore] connect individuals to private browsing sessions."²⁷⁰ But the document notes that "the promise not to do this is effectively already being applied," and the author of this email explained at his deposition that such joining does not occur.²⁷¹ This email also does not state that

²⁶⁸ Douglas C. Schmidt, et al., "Google data collection," Vanderbilt University, <https://digitalcontentnext.org/wpcontent/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf> (Aug. 15, 2018).

²⁶⁹ *Id.* at 21-22 ("The experiment presented below assessed whether Google can connect such identifiers (and hence information associated with them) with a user's personal information. This experiment involved the following ordered steps: 1. Opened a new (no saved cookies, e.g. Private or Incognito) browser session (Chrome or other), 2. Visited a 3rd-party website that used Google's DoubleClick ad network, 3. Visited the website of a widely used Google service (Gmail in this case), 4. Signed in to Gmail.").

²⁷⁰ Schneier ¶ 155.

²⁷¹ McClelland Feb. 18, 2022 Tr. 80:12-81:11.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

such joining would be possible by applying Plaintiffs' proposed methods for joining the data, nor does it state that Incognito sessions would be joined to specific users' identities.

- b. GOOG-CABR-05270014: Mr. Schneier cites a comment in an email chain stating

[REDACTED] "In light of this email, Mr. Schneier concludes that "[i]n other words, Google can connect individuals to private browsing sessions." Thus, Mr. Schneier states that the theoretical possibility that private browsing sessions can be joined with signed-in browsing is a sufficient condition for linking private browsing mode sessions with an individual user's identity. But as discussed above at [§ III.J](#), this incorrectly assumes a one-to-one connection between devices and users that does not account for shared devices or align with Mr. Schneier's own published research concluding that "households are not units; devices are not personal; [and] the purchaser of a product is not its only user."²⁷²

193. Finally, Mr. Schneier's claim that data is difficult to anonymize and easy to de-anonymize is incorrect as to the Data At Issue because, for the reasons discussed *supra* [§ III.A.3](#), [§ III.F.3](#), Google's policies, procedures, and technical solutions designed to anonymize the Data At Issue and prevent its de-anonymization have been effective.

C. Opinion 13: Mr. Schneier's Assertion That Google Has Not Taken Steps To Ensure That A User's Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User's Signed-Out Activity With Any Signed-In Data Is Incorrect

194. According to Mr. Schneier:

Google has not taken steps to ensure that a user's choice to sign out of a Google account will prevent Google from associating the user's signed-out activity with any signed-in data. The cookies that Google collects "span signed in and signed out sessions," allowing Google to "connect the dots even if [it] can't write data to a person's account." And even if Google is not building user profiles across signed-in and signed-out data, Google's decision to collect and log this data creates the potential for data to be joined in this way. For example, Google's storage of unique identifiers and IP addresses together in logs introduces a risk

²⁷² K. Levy and B. Schneier, "Privacy threats in intimate relationships," *Journal of Cybersecurity*, <https://academic.oup.com/cybersecurity/article/6/1/tyaa006/5849222> (2020), at 10.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

that data from a users' private browsing will be joined with a user's signed-in data.²⁷³

195. As explained further below, Mr. Schneier's assertion is incorrect for the following reasons:

- ❖ **First**, the documents that Mr. Schneier cites do not support his conclusion that Google "has not taken steps to ensure that a user's choice to sign out of a Google account will prevent Google from associating the user's signed-out activity with any signed-in data" and in many cases they show extensive steps to maintain strict separation between signed-in and signed-out data. *See infra* [§ IV.C.1.](#)
- ❖ **Second**, additional documents and testimony produced in this case further demonstrate the steps that Google has taken to ensure that a user's choice to sign out of a Google account will prevent Google from associating the user's signed-out activity with any signed-in data. *See infra* [§ IV.C.2.](#)
- ❖ **Third**, Google's policy restrictions and pseudonymization procedures closely align with best practices for research involving user data. *See infra* [§ IV.C.3.](#)
- ❖ **Fourth**, Google's security practices closely align with best practices in the network security industry. *See infra* [§ IV.C.4.](#)

1. The Documents Mr. Schneier Cites Do Not Support His Conclusion

196. In my opinion, none of the documents that Mr. Schneier cites supports his conclusion that Google "has not taken steps to ensure that a user's choice to sign out of a Google account will prevent Google from associating the user's signed-out activity with any signed-in data."

²⁷³ Schneier ¶ 205 (citing GOOG-BRWN-00060463; GOOG-CABR-00358713; GOOG-BRWN-00386570; GOOG-BRWN-00613801; GOOG-BRWN-00386402; GOOG-CABR-00799341).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

197. GOOG-BRWN-00060463 is a draft of potential issues and discussion points identified by a Google employee, and the discussion Mr. Schneier quotes is not directed to Incognito mode. Indeed, the document undermines Mr. Schneier's opinion because it states that the use of Incognito mode is an effective way to prevent joining of signed-in and signed-out browsing data, as it notes that [REDACTED]

[REDACTED]²⁷⁴

198. GOOG-CABR-00358713 identifies a potential risk of joining Incognito data with signed-in data, but (i) does not say that any such joining takes place; and (ii) notes that [REDACTED]

[REDACTED]²⁷⁵ As discussed above (*supra* [§ III.B](#)), allowing Chrome to “tell Google (or all sites) that a user went Incognito”²⁷⁶ would conflict with the W3C Guidelines recommending that the use of private browsing mode should not be detectable to websites. This discussion confirms that Google decided not to pursue solutions that would require the Chrome browser to send a signal to Google or third party websites that the browser was in incognito mode.

199. GOOG-BRWN-00386570 is an email discussing potential implementation of Incognito mode in the Google Search App (*i.e.*, a non-Browser app that is distinct from Chrome), and an email in the thread notes that “aside from the potential linking to the user's GAIA cookie ID in regular [mode] IF the IP is static, there is no way to link that signed-out/incognito zwieback cookie to Gaia.”²⁷⁷ Again, this email discusses potential

²⁷⁴ GOOG-BRWN-00060463, at -463 (emphasis added).

²⁷⁵ GOOG-CABR-00358713, at -713.

²⁷⁶ *Id.*

²⁷⁷ GOOG-BRWN-00386570, at -570.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

joinability *risks* related to a non-browser app (regarding search activity), rather than actual joining of user browsing data by Google. And as discussed above (*supra* § III.G.1.), static IPs are exceedingly rare, so the risk identified in this email does not appear to be substantial.

200. GOOG-BRWN-00613801 is a comment notification showing a discussion between Google employees regarding joining of data from a private browsing session with a “Google account . . . if [users] sign into it from incognito mode.” Since Plaintiffs’ proposed classes consist of users who are *not* signed into a Google account while using private browsing mode, this type of joining is not relevant to the Data At Issue.²⁷⁸ This discussion also notes that the proposal to send a signal to prohibit collection of data from an Incognito mode session raises: “concerns . . . about altering the behavior of Google sites during the session because it makes incognito in Chrome and other browsers less consistent and has the risk of actually decreasing privacy if some use-cases are broken in incognito mode and therefore force users to use regular mode.” Making “incognito in Chrome and [private browsing mode on] other browsers less consistent” would increase the risk that a third party website could detect the use of Incognito mode, which would violate the aforementioned W3C principles.

201. GOOG-BRWN-00386402 is an email where a Google employee notes that [REDACTED] Like GOOG-BRWN-00386570, this is a discussion of potential risks (regarding activity on Google websites)²⁷⁹ that

²⁷⁸ Complaint ¶ 192.

²⁷⁹ A Zwieback identifier is used for activity on Google “Owned and Operated” (“O&O”) properties like searches on google.com. See GOOG-CABR-00543864, at -958 [REDACTED]

[REDACTED] GOOG-BRWN-00078348. The use of Google owned and operated properties like google.com is outside of the scope of Plaintiffs’ proposed classes, which are limited to users “who accessed a non-Google website.” Complaint ¶ 192.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

affirmatively states that Google does not actually join such data. And as discussed above (*supra* [§ III.G.1.](#)), static IPs are exceedingly rare, so the risk identified in this email does not appear to be substantial.

202. These documents do not support Mr. Schneier’s conclusion that Google “has not taken steps to ensure that a user’s choice to sign out of a Google account will prevent Google from associating the user’s signed-out activity with any signed-in data.”²⁸⁰ Indeed, none of the documents cited by Mr. Schneier (or any other documents or testimony I have reviewed) state that joining of signed-in and signed-out user data actually takes place, which indicates that Google’s myriad policies and technical restrictions (discussed further *infra* [§ IV.C.2.](#)) have been effective in preventing *actual joining* of signed-in and signed-out data. In my opinion, Google’s implementation of (and adherence to) policies and technical restrictions to prevent joining of signed-out and signed-in data shows that Google *has* “taken steps” to segregate signed-in from signed-out data. And the documents Mr. Schneier cites indicate that Google employees investigate joinability risks regularly and propose steps for mitigating those risks. This process of self-evaluation, identification of risks, and introduction of mitigations is consistent with a proactive compliance regime designed to prevent the joining of signed-in and signed-out data, rather than evidence that such a regime has not been put into place. As discussed further below, additional documents Google has produced in this case also support this conclusion.

2. Additional Documents And Testimony Contradict Mr. Schneier’s Claim That Google Has Not Undertaken Steps To Prevent Joining of Signed-Out And Signed-In Data

203. The efforts Google has undertaken to prevent joining of signed-out and signed-in browsing data are further evidenced by additional documents produced in this action, including the following examples:

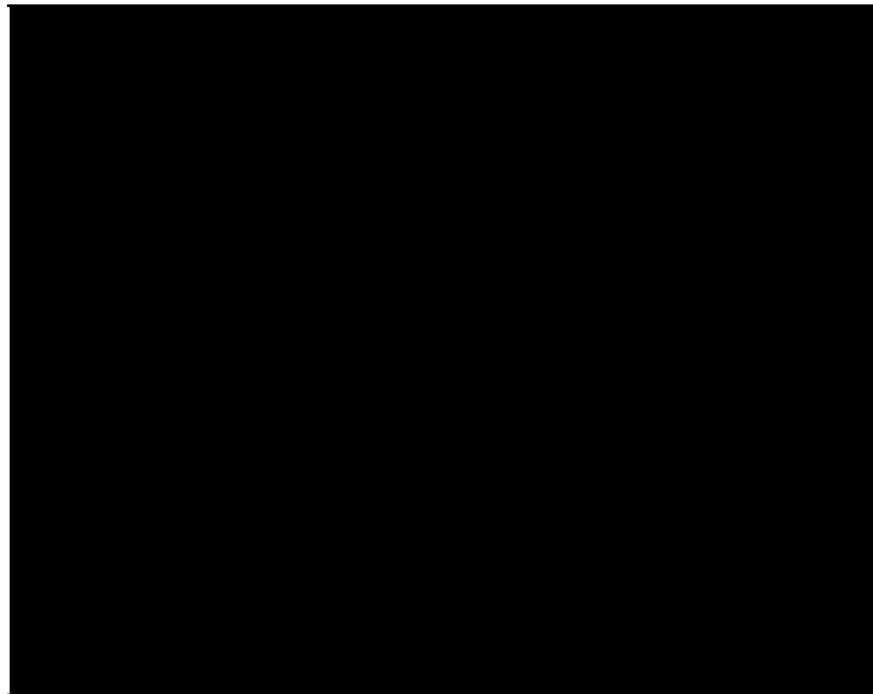
²⁸⁰ Schneier ¶ 205.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- a. GOOG-CABR-04716372, at -375 [REDACTED]
- [REDACTED] This design document from 2016 indicates that identifying and mitigating factors that may increase the risk of joining of signed-out and signed-in logs was and is a foundational principle in the design of Google's logging infrastructure.

- b. GOOG-BRWN-00426550, at -551 [REDACTED]
- [REDACTED] at -553
- [REDACTED] at -555
- [REDACTED] This design document shows that Google has implemented technical processes to prevent the joining of logged-in and logged-out activity.

- c. GOOG-CABR-05461707, at -707-08:



This retention policy further illustrates steps that Google has taken to mitigate the risk of joining of signed-in and signed-out browsing activity via “scrubbing” of “cookies and any other stable identifiers [in unauthenticated logs]. . . to lower . . . re-identification risk.”

- d. GOOG-CABR-00063770, at -770-71 [REDACTED]
- [REDACTED]

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

██████████ This logging policy shows additional restrictions that Google has implemented in order to maintain separation between authenticated (i.e., “Gaia keyed”) and unauthenticated (i.e., “anonymous”) user data.

e. GOOG-BRWN-00027368, at -404:



This slide (i) underscores that preventing the joining of signed-in and signed-out user data is a core privacy requirement; and (ii) shows how Google uses encryption in ██████████ to prevent such joining.

f. *Id.* at -406:



This slide shows how ██████████ uses encryption to further restrict joining of authenticated and unauthenticated data.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

204. I have also reviewed deposition testimony from a third-party ex-Google product manager (whose legal fees associated with the deposition were paid by Plaintiffs' counsel²⁸¹) that further confirms that Google does not join private and non-private browsing mode information to create "cradle-to-grave" profiles.²⁸² Additional testimony from other Google witnesses confirms that Google does not join signed-in and signed-out browsing mode information.²⁸³

205. Assessments by independent auditors also describe the extensive steps Google has undertaken to implement privacy controls related to user data.²⁸⁴

3. Google's Policy Restrictions And Pseudonymization Procedures Closely Align With Best Practices For Research Involving User Data

206. Throughout my professional career as an academic, and in particular during the last handful of years that I have been working on efficient, privacy-preserving systems in the

²⁸¹ See McClelland Tr. 228:23-229:3 ("[Q.] [D]o you understand Mr. Bailey to be charging you in connection with his representation of you at today's deposition? A. No, the costs are being covered by the plaintiffs.").

²⁸² See McClelland Tr. 290:10-24 ("Q. Specifically[,] [is] the allegation that Google creates a cradle to grave profile that [combines] log[ged] out private browsing data with non-private browsing data [accurate]? A. My understanding and my experience at Google would indicate that that is not the case. From memory, tracking data after [REDACTED] is discarded, it's no longer useful, so the longest profile is based upon [REDACTED]' worth of data. Signed-in data is never -- there is no attempt to link signed in -- authenticate[d] user data with non-signed in data and certainly Incognito sessions were not joined with regular sessions.").

²⁸³ See Berntson June 16, 2021 Tr. 199:1-8 ("Google has a set of, as -- as you noted, policies that are meant to prevent reidentifiability, prevent joining of sort of sensitive IDs in terms of, say, signed in and signed out. So the vast majority of the way all of our systems work, we maintain a very strict separation."); Berntson Mar. 18, 2022 Tr. 105:10-12 ("I'm not aware of an instance where [REDACTED] is joining IDs together for the purpose of serving personalized ads."); Jun Mar. 1, 2022 Tr. 146:10-19 ("[O]ne of the privacy principle[s] I learned during the project was we wanted to honor civilians' privacy and wanted to make sure it's not easy to map from device ID or biscotti ID to GAIA ID because GAIA ID might be linked to Gmail account which Google might be asked to retrieve the Gmail account contents through any government order, something like that. So vice versa. So we wanted to prevent such ID joinability to protect civilians' privacy.").

²⁸⁴ See, e.g., GOOG-BRWN-00526782; GOOG-BRWN-00468598; GOOG-BRWN-00041778; GOOG-BRWN-00468530.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

context of three National Science Foundation (NSF) grants, I have been involved in a number of research efforts that require collection of data related to human subjects.²⁸⁵ Both the National Institute of Health (NIH) and NSF have policies that researchers need to follow to protect the privacy rights of human subjects.²⁸⁶ These regulations make grantee institutions responsible for setting up "Institutional Review Boards" (IRBs) to review research protocols and designs and ensure the protection of the rights of human subjects including privacy rights. IRBs review research protocols and designs and are charged with approving (or not) exceptions to ensure the protection of the rights of human subjects including privacy rights in research institutions per NSF/NIH regulations. Researchers collecting and working with sensitive data need to adhere to university privacy policies. *See, e.g.*, the University of Southern California's (USC's) policy²⁸⁷ and Stanford University's policy.²⁸⁸

207. Like these IRBs, Google used an "Internal Privacy Policy Team" that (i) maintains privacy, data access, and other associated policies; and (ii) maintains control over those policies.²⁸⁹

²⁸⁵ National Science Foundation, SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet, Award# (USC): 1956435; National Science Foundation, CNS Core: Medium: Collaborative Research: Privacy-Preserving Mobile Crowdsourcing, Award# (USC): 1901488; National Science Foundation, NeTS: Spectrum Sharing Systems for Wireless Networks: Performance and Privacy Challenges.

²⁸⁶ National Science Foundation, "Human Subjects," <https://www.nsf.gov/bfa/dias/policy/human.jsp> (last visited June 3, 2022); National Institutes of Health, "Protecting Sensitive Data and Information Used in Research," https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.12_protecting_sensitive_data_and_information_used_in_research.htm (updated Dec. 2021).

²⁸⁷ Office for the Protection of Research Subjects, USC, <https://oprs.usc.edu/policies/> (last visited June 3, 2022); Office for the Protection of Research Subjects, USC, "Chapter 10: Privacy, Confidentiality and HIPAA," <https://oprs.usc.edu/policies/privacy-confidentiality-and-hipaa/> (last visited June 3, 2022).

²⁸⁸ Research Compliance Office, Stanford University, HRPP Policy Manual.pdf, <https://researchcompliance.stanford.edu/panels/hs/policies> (last visited June 3, 2022).

²⁸⁹ GOOG-BRWN-00200355, at -361-65.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

208. To protect sensitive data, researchers are required to, for example, maintain systems where (i) “[t]he re-identification algorithm, code, or pseudonym is maintained in a separate system, with the appropriate controls in place to prevent unauthorized access to re-identification information”; and (ii) “[a]ll of the identifying PII fields can be removed, and the patient ID numbers can be obscured using pseudo-random data that is associated with a cross-reference table located in a separate system.”²⁹⁰ Similarly, Google separates the so called [REDACTED] which handles PII from the rest of the services: “Because we want [REDACTED] to be the only service in Display Ads to have raw access to both Gaia and Biscotti identifier spaces, [REDACTED] has to run under a separate MDB group from those running other existing services.”²⁹¹

209. To protect sensitive data, researchers are required to “Limit access to personally identifiable information using the principle of strict need-to-know.”²⁹² Similarly, Google’s user data access policy²⁹³ dictates that “[a]uthorization is granted only for a specific purpose” and “[a]uthorized access to User Data for one purpose does not mean it can be used for another purpose.”

210. Encryption of data is also required in the research context by established federal entities and established educational research institutes. For example, NIH policy says “If portable electronic devices must be used, they should be encrypted to safeguard data and

²⁹⁰ National Institute of Standards and Technology, “Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)” <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-122.pdf> (Apr. 2010).

²⁹¹ GOOG-CABR-04715843, at -876; *see also* GOOG-CABR-00058926, at -963 (“Gaia encrypted with [REDACTED] . . . Biscotti encrypted with [REDACTED] . . .”).

²⁹² Office for the Protection of Research Subjects, USC, “Chapter 10: Privacy, Confidentiality and HIPAA,” <https://opr.s.usc.edu/policies/privacy-confidentiality-and-hipaa/> (last visited June 3, 2022).

²⁹³ GOOG-CABR-05455683, at -683.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

information.”²⁹⁴ USC’s policy further specifies that researchers should “[e]ncrypt data” and use “[e]ncryption for transmission.”²⁹⁵ Google’s users data access policy similarly states that “[u]ser Data must be strongly encrypted at rest and in transit.”²⁹⁶

211. Anonymization of data is strongly encouraged and anonymization policies are thoroughly tested. For example, Stanford’s policy states that “[the IRB] evaluates the proposed anonymizing techniques, (e.g., de-identification, coding), storage plans, access restrictions, data security methods (e.g., encryption) and other relevant factors in making its final determination concerning the appropriateness and adequacy of confidentiality protections,”²⁹⁷ and the corresponding Google’s policy states that “[a]ll User Data anonymization schemes should be approved by the Privacy Working Group’s Anonymization Team [because] [a]nonymizing data is difficult to get right.”²⁹⁸

4. Google’s Security Practices Closely Align With Best Practices In The Network Security Industry

212. As discussed further below, it is my opinion that Google’s information security practices regarding the Data At Issue are aligned with best practices in the network security industry.

213. There are a number of organizations that promulgate standards prescribing best practices for information security. For example, the Center for Internet Security (CIS) issues periodic “Critical Security Controls” guidelines that have been cited by the California Attorney

²⁹⁴ National Institute of Health, “NIH Grants Policy Statement,” https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.12_protecting_sensitive_data_and_information_used_in_research.htm (updated Dec. 2021).

²⁹⁵ Office for the Protection of Research Subjects, USC, “Chapter 10: Privacy, Confidentiality and HIPAA,” <https://oprs.usc.edu/policies/privacy-confidentiality-and-hipaa/> (last visited June 3, 2022).

²⁹⁶ GOOG-CABR-05455683, at -684.

²⁹⁷ Research Compliance Office, Stanford Univ., HRPP Policy Manual.pdf, <https://researchcompliance.stanford.edu/panels/hs/policies> (last visited June 2022), at 144.

²⁹⁸ GOOG-CABR-05455683, at -684.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

General as “security measures with a high payoff . . . [that are] the priority actions that should be taken as the starting point of a comprehensive program to provide reasonable security.”²⁹⁹ CIS recently released Version 8 of this standard, which consists of eighteen recommended security measures.³⁰⁰ “The CIS Controls were developed starting in 2008 by an international, grass-roots consortium bringing together companies, government agencies, institutions, and individuals from every part of the ecosystem (cyber analysts, vulnerability-finders, solution providers, users, consultants, policy-makers, executives, academia, auditors, etc.) who banded together to create, adopt, and support the CIS Controls.”³⁰¹ “The expert volunteers who develop the Controls apply their first-hand experience to develop the most effective actions for cyber defense.”³⁰²

214. A number of the CIS recommendations are particularly pertinent to the Data At Issue:

- a. CIS Critical Security Control 3 (“CSC No. 3”): Data Protection (“Develop processes and technical controls to identify, classify, securely handle, retain, and dispose of data.”)
- b. CIS Critical Security Control 6 (“CSC No. 6”): Access Control Management (“Use processes and tools to create, assign, manage, and revoke access credentials and privileges for user, administrator, and service accounts for enterprise assets and software.”)

²⁹⁹ Kamala D. Harris, “California Data Breach Report,” Cal. Dept. of Justice, <https://oag.ca.gov/breachreport2016> (2016), at 31.

³⁰⁰ Center For InternetSecurity, CIS Critical Security Controls Version 8 (“CIS Controls”), <https://www.cisecurity.org/controls/v8> (last visited June 6, 2022).

³⁰¹ Center for Internet Security, “CIS Critical Security Controls FAQ,” <https://www.cisecurity.org/controls/cis-controls-faq> (last visited June 3, 2022).

³⁰² *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

215. CSC No. 3 states that “[i]t is important for an enterprise to develop a data management process that includes a data management framework, data classification guidelines, and requirements for protection, handling, retention, and disposal of data.”³⁰³ This data management process should “address data sensitivity, data owner, handling of data, data retention limits, and disposal requirements, based on sensitivity and retention standards for the enterprise.”³⁰⁴ Retention policy enforcement entails “[r]etain[ing] data according to the enterprise’s data management process,” and “[d]ata retention must include both minimum and maximum timelines.”³⁰⁵ CSC No. 3 further recommends that sensitive data be encrypted “at rest” and “in transit.”³⁰⁶

216. In my opinion, Google’s policies and restrictions applicable to logs containing user data align with CSC No. 3 for the following reasons:

- a. Google’s policies applicable to logs containing user data clearly define the scope of logs to which the policies apply.³⁰⁷
- b. Google’s policies must be followed by all employees or other representatives.³⁰⁸

³⁰³ CIS Controls at 15.

³⁰⁴ *Id.*

³⁰⁵ *Id.*

³⁰⁶ *Id.*

³⁰⁷ *See, e.g.,* GOOG-CABR-00892455, at -455

³⁰⁸ *Id.* (“This policy applies to all full-time and part-time employees, the extended workforce, and other authorized representatives of Google and its subsidiaries.”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- c. Pursuant to Google's User Data Access Policy, "[a]ll User Data at Google has a Data Owner."³⁰⁹
- d. Google's User Data Access Policy also requires that "User Data must be **strongly** encrypted at rest and in transit."³¹⁰
- e. Google's retention policies applicable to user data establish retention periods applicable to user data in logs.³¹¹

217. CSC No. 6 states that "[t]here should be a process where privileges are granted and revoked for user accounts . . . ideally . . . based on enterprise role and need through role-based access."³¹² "Role-based access is a technique to define and manage access requirements for each account based on: need to know, least privilege, privacy requirements, and/or separation of duties."³¹³

218. In my opinion, Google's policies and restrictions applicable to logs containing user data align with CSC No. 6 for the following reasons:

- a. Google's policies establish "special rules" that, *inter alia*, govern "[a]uthorization and access control," "[a]ccessing data about users," "[r]e-identifying users from the data," "[c]onducting multi-day analysis," "[s]haring this data with others," and "[b]uilding tools that read or analyze this data."³¹⁴

³⁰⁹ GOOG-CABR-00073875, at -875.

³¹⁰ *Id.* at -876 (emphasis in original).

³¹¹ *See, e.g.*, GOOG-CABR-00051478; GOOG-CABR-04717190; GOOG-CABR-00057420.

³¹² CIS Controls at 24.

³¹³ *Id.*

³¹⁴ GOOG-CABR-00892455, at -455.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- b. These special rules are contained in a “User Data Access Policy,” and “User Data Access Authorization Guidelines,” that “apply to logs data.”³¹⁵
- c. Pursuant to Google’s User Data Access Policy, “User Data access is only granted for authorized, valid Google business purposes,” “[a]ll User Data access must be authorized by the appropriate User Data Delegate,” and “[a]uthorized access to User Data for one purpose does not mean it can be used for another purpose . . . [which means that you can] not access User Data for anything other than the original authorized purpose without seeking additional authorization.”³¹⁶

219. The analyses and opinions in this report are based on the results of my research, my review and analysis of the materials provided to me, my education and training, and my experience on related topics. As additional materials and information become available or if the scope of discovery or the causes of action change in any material way, I reserve the right to amend, supplement, or update my analysis and conclusions.

Respectfully submitted by,



Dr. Konstantinos Psounis
Date: June 7, 2022

³¹⁵ *Id.*

³¹⁶ GOOG-CABR-00073875, at -875.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

V. APPENDICES

A. Curriculum Vitae

EDUCATION

STANFORD UNIVERSITY Stanford, CA

Department of Electrical Engineering Jan. 1999 - Dec. 2002

Ph.D. degree in Electrical Engineering. Thesis title: “Probabilistic Methods for Web Caching and Performance Prediction of IP Networks and Web Farms”, supervised by Prof. Balaji Prabhakar.

STANFORD UNIVERSITY Stanford, CA

Department of Electrical Engineering Sep. 1997 - Jan. 1999

M.S. degree in Electrical Engineering. GPA: 4.0/4.0. (Actual GPA 4.05/4.0.)

NATIONAL TECHNICAL UNIVERSITY OF ATHENS Athens, Greece

Electrical and Computer Engineering Department Sep. 1992 - June 1997

Diploma in Electrical and Computer Engineering. GPA: 9.74/10.0. (Graduated ranking 1st in the class of '97.)

WORK IN ACADEMIA

UNIVERSITY OF SOUTHERN CALIFORNIA Los Angeles, CA

Professor Nov. 2017 - now

Associate Chair of ECE department Jan. 2019 - Aug 2019

Associate Professor May. 2009 - Nov. 2017

Assistant Professor Sep. 2003 - Apr. 2009

Electrical and Computer Engineering (ECE) and (jointly) Computer Science departments

Research interests: Modeling, performance analysis, algorithm design, and system implementation for efficient and privacy-preserving networked, distributed systems, including the Internet and the web, data centers, CDNs and cloud systems, WiFi/cellular and spectrum sharing systems, sensor and IoT systems, mobile ad hoc and delay tolerant networks, peer to peer systems, and autonomous vehicles systems.

STANFORD UNIVERSITY Stanford, CA

Visiting Associate Professor Aug. 2009 - Dec. 2009

Electrical Engineering department.

STANFORD UNIVERSITY Stanford, CA

Postdoctoral Research Fellow Jan. 2003 - Aug. 2003

Scheduling of Internet flows and multi-server systems.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

WORK IN INDUSTRY

ADANT TECHNOLOGIES

Consultant Jan. 2017 - May 2018
WiFi 802.11ax and smart antennas.

QUANTENNA COMMUNICATIONS

Consultant Jan. 2015 - 2016
Advanced MAC and PHY layer techniques applied to next generation wireless networks.

HONEYWELL

Consultant Sep. 2013 - Jun. 2014
Sensor-network based fire alarm systems.

SPACEMUX, INC.

Co-founder and CEO May 2013 - Dec. 2014
Increasing wireless bandwidth and speed tenfold.

CISCO SYSTEMS San Jose, CA

Consultant Sep. 2009 - Dec. 2009
Vehicular multi-technology wireless connectivity.

FINEGROUND NETWORKS INC. Cambell, CA

Technology Architect Sep. 2000 - June 2001
Accelerating web downloads using delta encoding.

TEACHING EXPERIENCE

- Co-creator and instructor for the USC graduate class EE597: “Wireless Networks”. Content: Introduction to current and next generation wireless networking technologies, detailed exploration of fundamental architectural and design principles used at all layers.
- Instructor for the USC graduate class EE503: “Probability for Electrical and Computer Engineering”. Content: Probability, discrete and continuous time Markov chains, basic queueing theory.
- Instructor for the USC undergrad class EE465: “Probabilistic Methods in Computer Systems Modelling”. Content: Probability, Markov chains, simulations.
- Creator and instructor for the USC graduate class EE650: “Advance Topics in Computer Networks: Mathematical tools for analyzing wired and wireless networks”. Content: Applications to networking problems of probability, queueing, Lyapunov functions, fluid limits, bipartite matchings, stable marriages, random walks on graphs, deterministic and stochastic optimization, statistical analysis, information theory, game theory.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

PHD STUDENT SUPERVISION

- Namo Asavisanu (Aug. 2021 - now)
- Sean Hackett (co-advised with Prof. Aleksandra Korolova, Aug. 2021 - now)
- Jiang Zhang (Aug. 2019 - now)
- Lillian Clark (co-advised with Prof. Bhaskar Krisnamachari, Sep. 2018 - now)
- Hang Qiu (co-advised with Prof. Ramesh Govindan, Jan. 2015 - June 2021, currently a Post Doctoral Fellow at Stanford University)
- Po-Han Huang (Sep. 2015 - Jan. 2020, currently at Facebook, California)
- Kaidong Wang (Sep. 2014 - Dec. 2019, currently at Qualcomm, California)
- Yonglong Zhang (Sep. 2013 - Dec. 2018, currently at Facebook, California)
- Matthew Clark (Sep. 2013 - Dec. 2017, currently at Aerospace Corporation, California)
- Weng Chon Ao (Sep. 2012 - Dec. 2017, currently at Qualcomm, California)
- Antonios Michaloliakos (Aug. 2010 - Sep. 2016, currently at Broadcom, California)
- Ranjan Pal (co-advised with Prof. Leana Golubchik, Jan. 2009 - Aug. 2014, currently Assistant Research Professor at the University of Michigan, Ann Arbor)
- Vlad Horia Balan (Sep. 2007 - Aug. 2013, currently at Google, California)
- Wei-Cherng Liao (Sep. 2004 - Dec. 2008, currently at MediaTek, Taiwan)
- Apoorva Jindal (Sep. 2003 - Dec. 2008, currently at Uber, California)
- Fragkiskos Papadopoulos (Sep. 2003 - Dec. 2007, currently Associate Professor at Cyprus University of Technology, Cyprus)
- Thrasyvoulos Spyropoulos (Sep. 2003 - Jun. 2006, currently Associate Professor at Eurecom, France)

PROPOSALS FUNDED

NSF SATC GRANT Oct. 2020 - Sep. 2025

National Science foundation (NSF) award under the Secure and Trustworthy Computing (SaTC) call.

Proposal title: SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet.

NSF CNS GRANT Oct. 2020 - Sep. 2023

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

National Science foundation (NSF) award under the Computer and Networked Systems (CNS) call.

Proposal title: CNS Core: Medium: Network-Enabled Cooperative Perception for Future Autonomous Vehicles.

NSF NETS GRANT Sep. 2019 - Aug. 2022

National Science foundation (NSF) award under the Networking Technology and Systems (NeTS) call.

Proposal Title: CNS Core: Medium: Collaborative Research: Privacy-Preserving Mobile Crowdsourcing.

CISCO SYSTEMS GRANT April 2019

Research grant from the Cisco University Research Program.

Proposal Title: Virtual and augmented reality over next generation WiFi.

CISCO SYSTEMS GRANT Dec. 2016

Research grant from the Cisco University Research Program. Proposal Title: Data-driven formal optimization of data centers.

NSF NETS GRANT Sep. 2016 - Aug. 2020

National Science foundation (NSF) award under the Networking Technology and Systems (NeTS) call.

Proposal Title: Spectrum Sharing Systems for Wireless Networks: Performance and Privacy Challenges.

ADANT TECHNOLOGIES GRANT June 2016

Research grant from Adant Technologies.

Proposal Title: Using reconfigurable antenna systems with WiFi communication devices.

HUAWEI GRANT May 2016

Research grant from Huawei.

Proposal Title: Addressing wireless bandwidth demand via asynchronously coordinated multi-cell deployments.

ADANT TECHNOLOGIES GRANT Dec. 2015

Research grant from Adant Technologies.

Proposal Title: Asynchronous coordination of WiFi transmitters equipped with smart antennas for enhanced spectral efficiency.

NSF EARS GRANT Sep. 2014 - Aug. 2019

National Science foundation (NSF) award under the Enhancing Access to the Radio Spectrum (EARS) crosscutting program.

Proposal Title: Future Wireless Broadband Access: Cross-Optimizing Hardware, Physical and Network Layers.

CISCO SYSTEMS GRANT May 2014

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Research grant from the Cisco University Research Program.
Proposal Title: Rateless encoded UDP for error-resilient wireless links.

ARMY RESEARCH LABORATORY (ARL) GRANT Sep. 2009 - Aug. 2014
CTA: Communications and Networking Academic Research Center.
Proposal Title: QUANTA: Quality of Information-Aware Networks for Tactical Applications.

DOCOMO LABS GRANT Sep. 2011 - 2013
Research support from the DoCoMo Labs, US. Proposal Title: MIMO systems with TDD

CISCO SYSTEMS GRANT Sep. 2011 - 2013
Research grant from the Cisco University Research Program. Proposal Title: Efficient airtime allocation in wireless networks.

MING HSIEH INSTITUTE (MHI) GRANT May 2011 - 2013
MHI grant to build a large scale software radio testbed and implement distributed MIMO, interference alignment and massive MIMO systems, as well as perform channel sounding and modelling.
Proposal Title: Large-Scale Software-Radio Testbed.

METRANS TRANSPORTATION CENTER GRANT Aug. 2009 - Aug. 2010
METRANS Transportation Center grant to conduct research on metropolitan transportation issues.
Proposal title: End-to-end performance in vehicular networks with an emphasis on safety and security applications.

CISCO SYSTEMS GRANT Sep. 2008
Research grant from the Cisco University Research Program.
Proposal Title: Neighborhood centric transport for home networking environments.

NSF NETS GRANT Aug. 2008 - Aug. 2011
National Science foundation (NSF) award under the Networking Technology and Systems (NeTS) call.
Proposal title: Contention-Awareness in Mesh Transport: Theory and Practice.

CISCO SYSTEMS GRANT Apr. 2008
Research grant from the Cisco University Research Program.
Proposal Title: TCP challenges in multi-hop wireless networks. From the networking workshop "The Future of TCP: Train-wreck or Evolution?".

NSF REU SITE GRANT Mar. 2008 - Mar. 2011
Grant to establish a National Science Foundation (NSF) Research Experiences for Under-graduates (REU) site within the Computer Science department at the Viterbi School of Engineering.
Proposal Title: Coordination, Communication, Autonomy: Principles and Technologies.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

VSOE INNOVATIVE RESEARCH FUND GRANT Dec. 2007 - Dec. 2008

Fund to initiate a Viterbi School of Engineering (VSoE) invited workshop on Wireless Networks. Proposal title: Establishing a New USC Invited Workshop on Theory and Practice in Wireless Networks.

METRANS TRANSPORTATION CENTER GRANT Oct. 2007 - Dec. 2008

METRANS Transportation Center grant to conduct research on metropolitan transportation issues.

Proposal title: Efficient Routing for Safety Applications in Vehicular Networks.

NSF NETS GRANT Aug. 2005 - Aug. 2008

National Science foundation (NSF) award under the Networking Technology and Systems (NeTS) call.

Proposal title: Efficient Routing in Delay Tolerant Networking.

ZUMBERGE FACULTY RESEARCH AND INNOVATION GRANT July 2005 - June 2006

The James H. Zumberge faculty research and innovation award is granted to a selected number of Professors at the University of Southern California. Proposal title: Routing in Intermittently Connected Mobile Networks.

CHARLES LEE POWELL SCHOLARSHIP GRANT Dec. 2003 - Dec. 2004

The Charles Lee Powell grant is granted to a selected number of Assistant Professors at the University of Southern California.

AWARDS

ACM DISTINGUISHED MEMBER Nov. 2019

The ACM Distinguished Member program recognizes up to 10 percent of ACM worldwide membership with at least 15 years of professional experience who have achieved significant accomplishments or have made a significant impact on the computing field.

IEEE FELLOW Jan. 2018

The IEEE Grade of Fellow is conferred by the IEEE Board of Directors upon a person with an outstanding record of accomplishments in any of the IEEE fields of interest. The total number selected in any one year cannot exceed one-tenth of one-percent of the total voting membership. IEEE Fellow is the highest grade of membership and is recognized by the technical community as a prestigious honor and an important career achievement.

DISTINGUISHED MEMBER OF 2018 IEEE INFOCOM TPC AWARD 2018

The IEEE Communications Society awards annually a select number of TPC members of its flagship conference IEEE Infocom with a Distinguished Member award.

DISTINGUISHED MEMBER OF 2016 IEEE INFOCOM TPC AWARD 2016

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

The IEEE Communications Society awards annually a select number of TPC members of its flagship conference IEEE Infocom with a Distinguished Member award.

ACM NOTABLE ARTICLE IN COMPUTING - BEST OF 2013 2014

Selection of paper “Modelling BitTorrent-like systems with many classes of users”, W.-C. Liao, F. Papadopoulos, K. Psounis, and C. Psomas, ACM Transactions on Modelling and Computer Simulation, Vol. 23, Issue 2, Article No. 13, May 2013.

MEPC BUSINESS PLAN COMPETITION - 2ND PLACE 2013

Presentation of SpaceMUX Inc., a USC spinoff startup specializing in advanced physical layer techniques applied to next generation wireless networks.

ACM SENIOR MEMBER AWARD Jan. 2009

The Senior Member grade recognizes those ACM members with at least 10 years of professional experience and 5 years of continuous professional membership who have demonstrated performance that sets them apart from their peers.

IEEE SENIOR MEMBER AWARD Nov. 2008

Qualifications for this distinction are at least ten years of professional practice and five years of significant performance as demonstrated by substantial engineering responsibility or achievement, publication of engineering and technical papers, books or inventions, and the development and teaching of engineering courses.

FUTURE OF TCP BEST PRESENTATION AWARD Apr. 2008

“Best and Most Compelling Presentation and Demonstration Award” at the networking workshop “The Future of TCP: Train-wreck or Evolution?” held at Stanford University and sponsored by Cisco Systems.

ZUMBERGE FACULTY RESEARCH AND INNOVATION AWARD July 2005

The James H. Zumberge faculty research and innovation award is granted to a selected number of Professors at the University of Southern California.

CHARLES LEE POWELL SCHOLARSHIP AWARD Dec. 2003

The Charles Lee Powell award is granted to a selected number of Assistant Professors at the University of Southern California.

ILLEANA AND ERIC BENHAMOU STANFORD GRADUATE FELLOWSHIP 1997 - 2002

Fellowship is awarded for four years to a very select number of PhD students based on academic merit.

BEST-STUDENT NATIONAL TECHNICAL UNIVERSITY OF ATHENS AWARD 1997

Awarded yearly to the student that graduates with the highest GPA across all departments of National Technical University of Athens.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

OTHER GRADUATE STUDIES AWARDS 1997 - 1998

Regent's Fellowship by University of California Berkeley, Charles Lee Powell Foundation Graduate Fellowship by Caltech, Gordon Y. S. Wu Fellowship in Engineering by Princeton University, Sage Fellowship by Cornell University.

PUBLICATIONS**REFEREED JOURNALS**

1. L. Clark, J. Galante, B. Krishnamachari, and K. Psounis. "A Queue-Stabilizing Framework for Networked Multi-Robot Exploration", *IEEE Robotics and Automation Letters*, February 2021.
2. W. Chon Ao, P. Huang and K. Psounis. "Joint Workload Distribution and Capacity Augmentation in Hybrid Datacenter Networks", *IEEE/ACM Transactions on Net-working*, October 2020.
3. R. Pal, K. Psounis, J. Crowcroft, F. Kelly, P. Hui, J. Kelly, A. Chatterjee, L. Golubchik, and S. Tarkoma. "When Are Cyber Blackouts in Modern Service Networks Likely? A Network Oblivious Theory On Cyber (Re)Insurance Feasibility", *ACM Transactions on Management Information Systems*, Article No.: 5, June 2020.
4. M. Clark and K. Psounis. "Optimizing Primary User Privacy in Spectrum Sharing Systems", *IEEE/ACM Transactions on Networking*, Vol. 28, Issue 2, April 2020.
5. W. Chon Ao and K. Psounis. "Resource-constrained Replication Strategies for Hierarchical and Heterogeneous Tasks", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 31, Issue 4, June 2020.
6. K. Wang and K. Psounis. "Efficient scheduling and resource allocation in 802.11ax multi-user transmissions", *Computer Communications, Elsevier*, Vol. 152, pp.171-186, February 2020.
7. R. Pal, L. Golubchik, K. Psounis and T. Bandyopadhyay. "On Robust Estimates of Correlated Risk in Cyber-Insured IT Firms: A First Look at Optimal AI-Based Estimates under Small Data", *ACM Transactions on Management Information Systems*, Article No.: 9, October 2019.
8. Y. Zhang and K. Psounis. "Efficient Indoor Localization via Switched-beam Antennas", *IEEE Transactions on Mobile Computing*, June 2019.
9. P. Huang and K. Psounis, "Optimal Backhauling for Dense Small-Cell Deployments Using mmWave Links", *Computer Communications Journal, Elsevier*, Vol: 139, April 2019.
10. R. Pal, L. Golubchik, K. Psounis, and P. Hui, "Security Pricing as Enabler of Cyber-Insurance: A First Look at Differentiated Pricing Markets", *IEEE Transactions on Dependable and Secure Computing*, Vol. 16, Issue 2, March-April 2019.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

11. W. Chon Ao and K. Psounis. "Data-locality-aware User Grouping in Cloud Radio Access Networks", *IEEE Transactions on Wireless Communications*, Vol: 17, Issue: 11, Nov. 2018.
12. Y. Zhang and K. Psounis. "Consistently High MIMO Rates via Switched-beam Antennas", *IEEE/ACM Transactions on Networking*, Vol: 26, Issue: 5, Oct. 2018.
13. W. Chon Ao and K. Psounis. "Fast Content Delivery via Distributed Caching and Small Cell Cooperation", *IEEE Transactions on Mobile Computing*, Vol: 17, Issue: 5, May 2018.
14. R. Pal, L. Golubchik, K. Psounis, "Improving Cyber-Security via Profitable Insurance Markets", *ACM SIGMETRICS Performance Evaluation Review*, Vol: 45, Issue 4, Mar. 2018.
15. M. Clark and K. Psounis. "Trading Utility for Privacy in Shared Spectrum Access Systems", *IEEE/ACM Transactions on Networking*, Vol. 26, Issue 1, February 2018.
16. A. Michaloliakos, W. C. Ao, K. Psounis and Y. Zhang. "Asynchronously Coordinated Multi-timescale beamforming architecture for multi-cell networks", *IEEE/ACM Transactions on Networking*, Vol. 26, Issue 1, February 2018.
17. W. Chon Ao and K. Psounis. "Approximation Algorithms for Online User Association in Multi-Tier Multi-Cell Mobile Networks", *IEEE/ACM Transactions on Networking*, Vol: 25, Issue: 4, August 2017.
18. M. Clark and K. Psounis. "Equal Interference Power Allocation for Efficient Shared Spectrum Resource Scheduling", *IEEE Transactions on Wireless Communications*, Vol: 16, Issue 1, January 2017.
19. A. Michaloliakos, R. Rogalin, Y. Zhang, K. Psounis and G. Caire. "Performance Modeling of Next-Generation WiFi Networks", *Computer Networks Journal*, Vol. 105, pp.150-165, August 2016.
20. R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloli-akos, V. Balan, and K. Psounis. "Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO", *IEEE Transactions on Wireless Communications*, Vol. 13, Issue 4, pp. 1815 - 1831, April 2014.
21. H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis and G. Caire. "AirSync: Enabling Distributed Multiuser MIMO with Full Spatial Multiplexing", *IEEE/ACM Transactions on Networking*, Vol. 21, Issue 6, pp. 1681 - 1695, December 2013.
22. A. Jindal and K. Psounis. "On the Efficiency of CSMA-CA Scheduling in Wireless Multihop Networks", *IEEE/ACM Transactions on Networking*, Vol. 21, Issue 5, pp. 1392 - 1406, October 2013.
23. W.-C. Liao, F. Papadopoulos, K. Psounis, and C. Psomas. "Modelling BitTorrent-like systems with many classes of users", *ACM Transactions on Modelling and Computer Simulation*, Vol. 23, Issue 2, Article No. 13, May 2013.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

24. A. Jindal, K. Psounis, and M. Liu, "CapEst: A Measurement-based Approach to Estimating Link Capacity in Wireless Networks", *IEEE Transactions on Mobile Computing*, Vol. 11, Iss. 12, pp. 2098–2108, May 2012.
25. S. Rangwala, A. Jindal, K.-Y. Jang, K. Psounis, and R. Govindan. "Neighborhood-centric congestion control for multi-hop wireless mesh networks", *IEEE/ACM Transactions on Networking*, Vol. 19, No. 6, pp. 1797–1810, December 2011.
26. W.-J. Hsu, T. Spyropoulos, K. Psounis and A. Helmy. "Modelling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks", *IEEE/ACM Transactions on Networking*, Vol. 17, Iss. 5, pp. 1564–1577, October 2009.
27. A. Jindal and K. Psounis. "The Achievable Rate Region of 802.11-Scheduled Multihop Networks", *IEEE/ACM Transactions on Networking*, Vol. 17, Iss. 4, pp. 1118–1131, August 2009.
28. A. Jindal, and K. Psounis. "Contention-Aware Performance Analysis of Mobility-Assisted Routing", *IEEE Transactions on Mobile Computing*, Vol. 8, No. 2, 145-161, February 2009.
29. T. Spyropoulos, K. Psounis, and C. Raghavendra. "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-copy Case", *IEEE/ACM Transactions on Networking*, Vol. 16, Iss. 1, pp. 77–90, February 2008.
30. T. Spyropoulos, K. Psounis, and C. Raghavendra. "Efficient Routing in Intermittently Connected Mobile Networks: The Single-copy Case", *IEEE/ACM Transactions on Networking*, Vol. 16, Iss. 1, pp. 63–76, February 2008.
31. F. Papadopoulos and K. Psounis. "Efficient Identification of Uncongested Internet Links for Topology Downscaling", *ACM SIGCOMM Computer Communication Review (CCR)*, Vol. 37, Issue 5, pp. 39–52, October 2007.
32. W.-C. Liao, F. Papadopoulos and K. Psounis. "Performance Analysis of BitTorrent-like Systems with Heterogeneous Users", *Performance Evaluation Journal*, Vol. 64, Issues 9–12, pp. 876-891, October 2007.
33. F. Papadopoulos, K. Psounis, and R. Govindan. "Performance Preserving Topological Downscaling of Internet-like Networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, special issue on "Sampling the Internet: Techniques and Applications", Vol. 24, No. 12, pp. 2313-2326, December 2006.
34. W.-C. Liao, F. Papadopoulos, and K. Psounis. "A Peer-to-peer Cooperation Enhancement Scheme and its Performance Analysis", *Journal of Communications (JCM)*, Vol. 1, No. 7, pp. 24–35, November/December 2006.
35. A. Jindal and K. Psounis. "Modelling Spatially Correlated Data in Sensor Networks", *ACM Transactions on Sensor Networks*, Vol. 2, Issue 4, pp. 466 - 499, November 2006.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

36. S. Rangwala, R. Gummandi, R. Govindan, and K. Psounis. “Interference-aware Fair Rate Control in Wireless Sensor Networks”, *ACM SIGCOMM Computer Communication Review (CCR)*, Vol. 36, Issue 4, pp. 63–74, October 2006.
37. W.-C. Liao, F. Papadopoulos, and K. Psounis. “An Efficient Algorithm for Resource Sharing in Peer-to-peer Networks”, *Lecture Notes in Computer Science, Springer*, Vol. 3976/2006, pp. 592–605, April 2006.
38. K. Psounis, P. Molinero Fernandez, B. Prabhakar, and F. Papadopoulos. “Systems with Multiple Servers under Heavy-tailed Workloads”, *Performance Evaluation Journal*, Vol. 62, Issue 1–4, pp. 456–474, October 2005.
39. R. Pan, K. Psounis, B. Prabhakar, and D. Wischik. “SHRiNK: A Method for Enabling Scaleable Performance Prediction and Efficient Network Simulation”, *IEEE/ACM Transactions on Networking*, Vol. 13, No. 5, pp. 975–988, October 2005.
40. J. Faruque, K. Psounis, and A. Helmy. “Analysis of Gradient-based Routing Protocols in Sensor Networks”, *Lecture Notes in Computer Science, Springer-Verlag*, Vol. 3560/2005, pp. 258–275, July 2005.
41. K. Psounis, A. Zhu, B. Prabhakar, and R. Motwani. “Modelling Correlations in Web-Traces and Implications for Designing Replacement Policies”, *Computer Networks Journal*, Vol. 45, No. 4, pp. 379–398, July 2004.
42. K. Psounis, R. Pan, B. Prabhakar, and D. Wischik. “The Scaling Hypothesis: Simplifying the Prediction of Network Performance Using Scaled-down Simulations”, *ACM SIGCOMM Computer Communication Reviews*, Vol. 33, No. 1, pp. 35–40, January 2003.
43. K. Psounis and B. Prabhakar. “Efficient Randomized Web-Cache Replacement Schemes Using Samples from Past Eviction-Times”, *IEEE/ACM Transactions on Networking*, Vol. 10, No. 4, pp. 441–454, August 2002.
44. K. Psounis, R. Pan, and B. Prabhakar. “An Approximate Fair Dropping Scheme for Variable Length Packets”, *IEEE Micro*, Vol. 21, No. 1, pp. 48–56, January/February 2001.
45. K. Psounis. “Active Networks, Applications, Security, Safety, and Architectures”, *IEEE Communications Surveys Magazine*, Vol. 2, No. 1, pp. 1–16, 1st quarter 1999.

CONFERENCE, PEER-REVIEWED, FULL-LENGTH PAPERS

1. J. Zhang, K. Psounis, H. Muhammad, and Z. Shafiq, “HARPO: Learning to Subvert Online Behavioral Advertising”, to appear in *NDSS*, 2022.
2. L. Clark, C. Andre, J. Galante, B. Krishnamachari, and K. Psounis, “TEAM: Trilateralization for Exploration and Mapping with Robotic Networks”, in *Proceedings of the 18th International Conference on Ubiquitous Robots*, July 2021.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

3. L. Clark, J. Galante, B. Krishnamachari, and K. Psounis, "A Queue-Stabilizing Framework for Networked Multi-Robot Exploration", in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, May 2021.
4. P.-H. Huang and K. Psounis, "Efficient User-Cell Association for 360 Video Streaming over Wireless Networks", in *Proceedings of IFIP Networking*, June 2020.
5. A. Petropulu, K. Psounis, and A. Al Hilli, "MIMO Radar Privacy Protection Through Gradient Enforcement in Shared Spectrum Scenarios", in *Proceedings of IEEE Inter-national Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Newark, NJ, November 2019.
6. E. Alimpertis, A. Markopoulou, C. T. Butts and K. Psounis, "City-Wide Signal Strength Maps: Prediction with Random Forests", in *Proceedings of WWW*, San Fransisco, CA, May 2019. (acceptance rate 15 percent)
7. A. Dimas, M. Clark, B. Li, K. Psounis and A. Petropulu, "On Radar Privacy in Shared Spectrum Scenarios", in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.
8. K. Wang and K. Psounis. "Scheduling and Resource Allocation in 802.11ax", in *Proceedings of IEEE Infocom*, Honolulu, Hi, April 2018. (acceptance rate 19.2 percent)
9. M. Clark and K. Psounis. "Achievable Privacy-Performance Tradeoffs for Spectrum Sharing with a Sensing Infrastructure", in *Proceedings of the 14th Annual Conference on Wireless On-Demand Network Systems and Services (IFIP WONS)*, 8 pages (no pp. avail.), Isola, France, February 2018.
10. Y. Zhang and K. Psounis. "Efficient MU-MIMO via Switched-beam Antennas", in *Proceedings of ACM MOBIHOC*, 10 pages (no pp. avail.), Madras, India, July 2017. (acceptance rate 17 percent)
11. M. Clark and K. Psounis. "Designing Sensor Networks to Protect Primary Users in Spectrum Access Systems", in *Proceedings of the 13th Annual Conference on Wireless On-Demand Network Systems and Services (IFIP/IEEE WONS)*, 8 pages (no pp. avail.), Jackson, WY, February 2017. (acceptance rate 30 percent)
12. P.-H. Huang and K. Psounis. "Efficient mmwave wireless backhauling for dense small-cell deployments", in *Proceedings of the 13th Annual Conference on Wireless On-Demand Network Systems and Services (IFIP/IEEE WONS)*, 8 pages (no pp. avail.), Jackson, WY, February 2017. (acceptance rate 30 percent)
13. W. Chon Ao and K. Psounis. "An Efficient Approximation Algorithm for Online Multi-Tier Multi-Cell User Association", in *Proceedings of ACM MOBIHOC*, 10 pages (no pp. avail.), Paderborn, Germany, July 2016. (acceptance rate 18.7 percent)
14. H. Qiu, K. Psounis, G. Caire, K. Chugg and K. Wang. "High-Rate WiFi Broadcasting in Crowded Scenarios via Lightweight Coordination of Multiple Access Points", in

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Proceedings of ACM MOBIHOC*, 10 pages (no pp. avail.), Paderborn, Germany, July 2016. (acceptance rate 18.7 percent)
15. M. Clark and K. Psounis. “Can the Privacy of Primary Networks in Shared Spectrum be Protected?”, in *Proceedings of IEEE INFOCOM*, 9 pages (no pp. avail.), San Francisco, April 2016. (acceptance rate 18.2 percent)
 16. G. Zois, A. Michaloliakos, K. Psounis, V. Vassalos and I. Mourtos. “Non-asymptotic performance bounds for downlink MU-MIMO scheduling”, in *Proceedings of the 12th Annual Conference on Wireless On-Demand Network Systems and Services (IFIP WONS)*, 8 pages (no pp. avail.), Italy, January 2016. (acceptance rate 30%)
 17. W. Chon Ao and K. Psounis. “Distributed Caching and Small Cell Cooperation for Fast Content Delivery”, in *Proceedings of ACM MOBIHOC*, pp. 127–136, Hangzhou, China, June 2015. (acceptance rate 14.8%)
 18. M. Clark and K. Psounis. “Efficient Resource Scheduling for a Secondary Network in Shared Spectrum”, in *Proceedings of IEEE INFOCOM*, pp. 1257–1265, Hong Kong, April 2015. (acceptance rate 19.0 percent)
 19. R. Pal, L. Golubchik, K. Psounis, and P. Hui. “Will Cyber-Insurance Improve Network Security? A Market Analysis”, in *Proceedings of IEEE INFOCOM*, pp. 235–243, Toronto, Canada, April 2014. (acceptance rate 19.4 percent)
 20. E. N. Ciftcioglu, A. Michaloliakos, K. Psounis, T. La Porta, and A. Yener. “Power Minimization with Quality-of-Information Outages”, in *Proceedings of the IEEE Wire-less Communications and Networking Conference (WCNC)*, pp. 1655–1660, Istanbul, Turkey, April 2014.
 21. R. Pal, L. Golubchik, K. Psounis, and P. Hui. “On A Way to Improve Cyber-Insurer Profits: When A Security Vendor Becomes the Cyber-Insurer”, in *Proceedings of IFIP NETWORKING*, 9 pages (no pp. avail.), New York, May 2013. (acceptance rate 26.2 percent)
 22. H. V. Balan, M. Segura, S. Deora, A. Michaloliakos, R. Rogalin, K. Psounis and G. Caire. “USC SDR, an easy-to-program, high data rate, real time software radio platform”, in *Proceedings of the ACM SIGCOMM workshop of Software Radio Implementation Forum (SRIF 2013)*, pp. 25–30, Hong Kong, China, August 2013.
 23. A. Michaloliakos, R. Rogalin, H. V. Balan, K. Psounis and G. Caire. “Efficient MAC for distributed multiuser MIMO systems”, in *Proceedings of the 10th Annual Conference on Wireless On-Demand Network Systems and Services (IFIP/IEEE WONS)*, pp. 52–59, Alberta, March 2013.
 24. H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis and G. Caire, “Achieving High Data Rates in a Distributed MIMO System”, in *Proceedings of ACM MOBICOM*, pp. 41–52, Istanbul, Turkey, August 2012. (acceptance rate 15.1 percent)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

25. M. Mongiovi, A. Singh, X. Yan, B. Zong, and K. Psounis, “Efficient multicasting for delay tolerant networks using graph indexing”, in *Proceedings of IEEE INFOCOM*, pp. 1386–1394, Orlando, Florida, March 2012. (acceptance rate 18.0 percent)
26. R. Pal, L. Golubchik, and K. Psounis. “Aegis: A Novel Cyber-Insurance Model”, in *Proceedings of the 2nd Conference on Decision and Game Theory for Security (GameSec 2011)*, pp. 131–150, College Park, Maryland, November 2011.
27. E. N. Ciftcioglu, A. Yener, R. Govindan, and K. Psounis. “Operational Information Content Sum Capacity: Formulation and Examples”, in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pp. 1–7, Chicago, July 2011.
28. K.-Y. Jang, K. Psounis, and R. Govindan. “Simple Yet Efficient, Transparent Airtime Allocation for TCP in Wireless Mesh Networks”, in *Proceedings of ACM CoNEXT*, article no. 28, 12 pages, Philadelphia, December 2010. (acceptance rate 19 percent)
29. A. Jindal and K. Psounis. “Making the Case for Random Access Scheduling in Wireless Multi-hop Networks”, in *Proceedings of IEEE INFOCOM*, (mini-conference), pp. 1–5, San Diego, California, March 2010. (acceptance rate 24 percent)
30. S. Rangwala, A. Jindal, K.-Y. Jang, K. Psounis, and R. Govindan. “Understanding Congestion Control in Multi-hop Wireless Mesh Networks”, in *Proceedings of ACM MOBICOM*, pp. 291–302, San Fransisco, California, September 2008. (acceptance rate 12 percent)
31. F. Papadopoulos and K. Psounis. “Scaling Properties of IEEE 802.11 Wireless Networks”, in *Proceedings of the 6th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 8 pages (no pp. avail.), Berlin, Germany, March 2008.
32. W.-C. Liao, F. Papadopoulos and K. Psounis. “Performance Analysis of BitTorrent-like Systems with Heterogeneous Users”, in *Proceedings of the 26th International Sym-posium on Computer Performance, Modeling, Measurements and Evaluation (IFIP Performance)*, pp. 876–891, Cologne, Germany, October 2007. (acceptance rate 23 percent)
33. W.-J. Hsu, T. Spyropoulos, K. Psounis and A. Helmy. “Modeling Time-variant User Mobility in Wireless Mobile Networks”, in *Proceedings of IEEE INFOCOM*, pp. 758–766, Anchorage , Alaska, USA, May 2007. (acceptance rate 18 percent)
34. A. Jindal, and K. Psounis. “Contention-Aware Analysis of Routing Schemes for Mobile Opportunistic Networks”, in *Proceedings of ACM MOBISYS, on the 1st International Workshop on Mobile Opportunistic Networking (MobiOpp)*, pp. 1–8, San Juan, Puerto Rico, June 2007.
35. F. Papadopoulos and K. Psounis. “Predicting the Performance of Mobile Ad hoc Networks Using Scaled-down Replicas”, in *Proceedings of IEEE International Conference on Communications (ICC)*, pp. 3928–3935, Glasgow, Scotland, June 2007.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

36. T. Spyropoulos, K. Psounis, and C. Raghavendra. "Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility", in *Proceedings of IEEE PERCOM, on the International Workshop on Intermittently Connected Mobile Ad hoc Networks (ICMAN)*, pp. 79–85, New York City, USA, March 2007.
37. A. Jindal and K. Psounis. "Fundamental Mobility Properties for Realistic Performance Analysis of Intermittently Connected Mobile Networks", in *Proceedings of IEEE PERCOM, on the International Workshop on Intermittently Connected Mobile Ad hoc Networks (ICMAN)*, pp. 59–64, New York City, USA, March 2007.
38. S. Rangwala, R. Gummandi, R. Govindan, and K. Psounis. "Interference-aware fair rate control in wireless sensor networks", in *Proceedings of ACM SIGCOMM*, pp. 63–74, Pisa, Italy, September 2006. (acceptance rate 12 percent)
39. T. Spyropoulos, K. Psounis, and C. Raghavendra, "Performance Analysis of Mobility-assisted Routing, in *Proceedings of ACM MOBIHOC*, pp. 49–60, Florence, Italy, May 2006. (acceptance rate 10 percent)
40. W.-C. Liao, F. Papadopoulos, and K. Psounis. "An Efficient Algorithm for Resource Sharing in Peer-to-peer Networks", in *Proceedings of IFIP Networking*, pp. 592–605, Coimbra, Portugal, May 2006. (acceptance rate 20 percent)
41. A. Jindal and K. Psounis. "Performance Analysis of Epidemic Routing under Contention", in *Proceedings of IWCMC*, pp. 539–544, Vancouver, Canada, July 2006.
42. K. Psounis, P. Molinero Fernandez, B. Prabhakar, and F. Papadopoulos. "Systems with Multiple Servers under Heavy-tailed Workloads", in *Proceedings of the 24th International Symposium on Computer Performance, Modeling, Measurements and Evaluation (IFIP Performance)*, pp. 456–474, Juan-les-Pins, France, October 2005. (acceptance rate 22 percent)
43. A. Jindal and K. Psounis. "Modeling Spatially-correlated Data of Sensor Networks with Irregular Topologies", in *Proceedings of IEEE SECON*, pp. 305–316, Santa Clara, California, USA, October 2005. (acceptance rate 27 percent)
44. T. Spyropoulos, K. Psounis, and C. Raghavendra. "Spary and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks", in *Proceedings of ACM SIGCOMM workshop on Delay Tolerant Networking (WDTN)*, pp. 252–259 Philadelphia, Philadelphia, USA, August 2005. (acceptance rate 22 percent)
45. J. Faruque, K. Psounis, and A. Helmy. "Analysis of Gradient-based Routing Protocols in Sensor Networks", in *Proceedings of IEEE/ACM DCOSS*, pp. 258–275, Marina Del Rey, California, USA, June 2005. (acceptance rate 28 percent)
46. K. Psounis, A. Ghosh, B. Prabhakar, and G. Wang. "SIFT: a Simple Algorithm for Trucking Elephant Flows and Taking Advantage of Power Laws", in *Proceedings of the 43rd Allerton Conference on Communication, Control, and Computing*, 10 pages (no pp. avail.), Urbana-Champaign, Illinois, USA, September 2005.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

47. F. Papadopoulos, K. Psounis, and R. Govindan. "Performance-Preserving Network Downscaling", in *Proceedings of the 38th Annual Simulation Symposium (ANSS)*, pp. 285–294, San Diego, California, April 2005.
48. A. Jindal and K. Psounis. "Modelling Spatially-correlated Sensor Network Data", in *Proceedings of IEEE SECON*, pp. 162–171, Santa Clara, California, USA, October 2004. (acceptance rate 19 percent)
49. T. Spyropoulos, K. Psounis, and C. Raghavendra. "Single-copy Routing in Intermittently Connected Mobile Networks", in *Proceedings of IEEE SECON*, pp. 235–244, Santa Clara, California, USA, October 2004. (acceptance rate 19 percent)
50. R. Pan, B. Prabhakar, K. Psounis, and D. Wischik. "SHRiNK: A Method for Scalable Performance Prediction and Efficient Network Simulation", in *Proceedings of IEEE INFOCOM*, Vol. 3, pp. 1943–1953, San Francisco, California, USA, April 2003. (acceptance rate 21 percent)
51. K. Psounis, R. Pan, B. Prabhakar, and D. Wischik. "The Scaling Hypothesis: Simplifying the Prediction of Network Performance Using Scaled-down Simulations", in *Proceedings of ACM HOTNETS*, pp. 35–40, Princeton, New Jersey, USA, October 2002.
52. R. Pan, B. Prabhakar, K. Psounis, and M. Sharma. "A Study of the Applicability of a Scaling Hypothesis", in *Proceedings of ASCC*, 6 pages (no pp. avail.), Singapore, Singapore, September 2002.
53. K. Psounis. "Class-based Delta Encoding: A Scalable Scheme for Caching Dynamic Web Content", in *Proceedings of IEEE ICDCS Workshops*, pp. 799 - 805, Vienna, Austria, July 2002.
54. K. Psounis and B. Prabhakar. "A Randomized Web-cache Replacement Scheme", in *Proceedings of IEEE INFOCOM*, Vol. 3, pp. 1407–1415, Anchorage, Alaska, USA, April 2001. (acceptance rate 23 percent)
55. R. Pan, B. Prabhakar, and K. Psounis. "CHOKe, A Stateless Active Queue Management Scheme for Approximating Fair Bandwidth Allocation", in *Proceedings of IEEE INFOCOM*, Vol. 2, pp. 942–951, Tel Aviv, Israel, March 2000. (acceptance rate 26 percent)
56. K. Psounis, R. Pan, and B. Prabhakar. "An Approximate Fair Dropping Scheme for Variable Length Packets", in *Proceedings of Hot Interconnects 8*, pp. 2–10, Stanford, California, USA, August 2000.
57. K. Psounis, B. Prabhakar, and D. Engler. "A Randomized Cache Replacement Scheme Approximating LRU", in *Proceedings of the 34th annual conference on Information Sciences and Systems*, 6 pages (no pp. avail.), Princeton, New Jersey, USA, March 2000.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

BOOK CHAPTERS

1. K. Psounis and M. Clark. Privacy in Spectrum Sharing Systems with Applications to Communications and Radar, In *Signal Processing for Joint Radar-Communications*, Wiley-IEEE Press, 2021.

INVITED JOURNALS

1. T. Spyropoulos, A. Jindal, and K. Psounis. “An Analytical Study of Fundamental Mobility Properties for Encounter-based Protocols”, *International Journal of Au-tonomous and Adaptive Communications Systems*, Vol. 1, Issue 1, pp. 440, July 2008.

INVITED CONFERENCE PAPERS

1. L. Clark, M. Clark, K. Psounis and P. Kairouz. “Privacy-utility trades in wireless data via optimization and learning”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, 10 pages (no pp. avail.), San Diego, California, USA, February 2019.
2. A. Dimas, B. Li, M. Clark, K. Psounis, A. Petropulu. “Spectrum Sharing Between Radar and Communication systems: Can the Privacy of the Radar be Preserved?”, in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, USA, October 2017.
3. A. Michaloliakos, W. Chon Ao and K. Psounis. “Joint user-beam selection for hybrid beamforming in asynchronously coordinated multi-cell networks”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, 10 pages (no pp. avail.), San Diego, California, USA, February 2016.
4. Y. Zhang, D. Bethanabhotla, T. Hao and K. Psounis. “Near-optimal user-cell association schemes for real-world networks”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, 10 pages (no pp. avail.), San Diego, California, USA, February 2015.
5. A. Jindal, K. Psounis, and M. Liu. “CapEst: Estimating wireless link capacity in multi-hop networks”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, 6 pages (no pp. avail.), San Diego, California, USA, February 2011.
6. D. Antonellis, A. Mansy, K. Psounis, and M. Ammar. “Real time, distributed network classification for routing protocol selection in mobile ad hoc networks”, in *Proceedings of the fourth international wireless Internet conference (WICON)*, 8 pages (no pp. avail.), Maui, Hawaii, November 2008.
7. Y. Wang, A. Ahmed, B. Krishnamachari, and K. Psounis. “IEEE 802.11p performance evaluation and protocol enhancement”, in *Proceedings of the IEEE International*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Conference on Vehicular Electronics and Safety, pp. 317–322, Columbus, Ohio, USA, September 2008.

8. A. Jindal and K. Psounis. “Achievable Rate Region and Optimality of Multi-hop Wireless 802.11-Scheduled Networks”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, 7 pages (no pp. avail.), San Diego, California, USA, January 2008.
9. F. Papadopoulos and K. Psounis. “Application of the many sources asymptotic in downscaling Internet-like networks”, in *Proceedings of the Information Theory and Applications Workshop (ITA)*, pp. 314–322, San Diego, California, USA, January 2007.
10. A. Jindal and K. Psounis. “Optimizing Multi-Copy Routing Schemes for Resource Constrained Intermittently Connected Mobile Networks”, in *Proceedings of the Forti-eth Asilomar Conference on Signals, Systems and Computers*, pp. 2142–2146, Pacific Grove, California, USA, October 2006.

CITATIONS

- Total citations: 12997
- h-index: 38

(source: google scholar, accessed: Fall 2021)

ISSUED PATENTS

- G. Caire, K. Psounis. Composite beamforming to coordinate concurrent WLAN links. Quantenna Communications, Inc.
- *US Patent No. 9,479,240*, issued Oct. 2016.
- K. Psounis, G. Caire, H. V. Balan. AirSync: enabling distributed multiuser MIMO with full multiplexing gain. USC.
- *US Patent No. 61,651,964*, issued Jan. 2015.
- K. Psounis and J. Jawahar. Method and System for Class-based Management of Dynamic Content in a Networked Environment. Cisco Systems, Inc.
- *US Patent No. 7,802,014*, issued Sep. 2010.
- R. Pan, B. Prabhakar and K. Psounis. A Stateless Active Queue Management Scheme for Approximating Fair Bandwidth Allocation. Stanford.
- *US Patent No. 7,324,442*, issued Jan. 2008.

SELECTED PROFESSIONAL SERVICE**INTERNATIONAL CONFERENCES – ORGANIZING/EXECUTIVE COMMITTEE**

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Steering Committee, IFIP/IEEE WONS, 2017 - now.
- General Chair, ACM SIGMETRICS, 2018.
- General Chair, IFIP/IEEE WONS, 2017.
- Program Chair, IFIP/IEEE WONS, 2016.
- Program Chair, IEEE DCOSS workshop on Wireless Sensor Networks (PWSN), 2014.
- Program Chair, ACM MOBICOM workshop on Challenged Networks (CHANTS), 2008.
- Workshop Chair, ACM SIGMETRICS 2008.
- Workshop Chair, USC Workshop on Theory and Practice in Wireless Networks, 2008.
- Publication Chair, ACM SIGMETRICS 2007.
- Panel Chair, ACM MOBIHOC, 2009.
- Panel Chair, IEEE CCW, 2008.

INTERNATIONAL CONFERENCES – TECHNICAL PROGRAM COMMITTEE

- IFIP/IEEE WONS 2013 - 2014, 2018, 2022
- IEEE INFOCOM 2005 - 2020.
- ACM SIGMETRICS 2008, 2014, 2015, 2017, 2020.
- ACM MOBIHOC, 2008 - 2010, 2017-2020.
- WiOpt 2016-2017.
- IEEE SECON 2007 - 2010.
- IFIP NETWORKING 2006 - 2010.
- ACM MOBICOM, 2009.
- IEEE ICNP 2009.
- IEEE WOWMOM workshop on Autonomic and Opportunistic Communications (AOC), 2008 -2009.
- IEEE ICDCS workshop on Delay Tolerant Mobile Networks (DTMN), 2008.
- ACM MOBISYS workshop on Mobile Opportunistic Networks (MOBIOPP), 2007.
- IEEE PERCOM workshop on Intermittently Connected Mobile Ad hoc Networks (IC-MAN), 2007.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

JOURNALS

- Editorial Board, IEEE/ACM Transactions on Networking (ToN), 2015 - 2020.
- Editorial Board, IEEE Transactions on Mobile Computing (TMC), 2009 - 2019.
- Editorial Board, Computer Networks Journal, Elsevier, 2009 - 2010.
- Editorial Board, International Journal of Autonomous and Adaptive Communications Systems (IJAACS), 2008.
- Reviewer of IEEE/ACM Transactions on Networking, IEEE Journal on Selected Areas in Communication, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Mobile Computing, ACM Transactions on Sensor networks, Elsevier Computer Networks Journal, Elsevier Performance Evaluation Journal, Elsevier Ad Hoc Networks Journal, Transportation Research Journal Part C, IEEE Transactions on Automatic Control.

GOVERNMENTAL AGENCIES

- NSF ML panel, 2020.
- NSF CAREER panel, 2019.
- NSF EARS meeting, 2016.
- NSF Future Internet Architecture Summit participant, 2009.
- NSF CRI panel member, 2008.
- NSF Wireless mobile workshop participant, 2007.
- NSF NeTS-NOSS panel member, 2005.
- Reviewer of NSF NeTS proposals.

PROFESSIONAL ASSOCIATIONS

- Institute of Electrical and Electronic Engineers (IEEE):
IEEE Fellow, 2018 - now.
Senior Member, 2008 - 2017.
Member, 1998 - 2008.
- Association for Computing Machinery (ACM):
ACM Distinguished Member, 2019 - now.
Senior Member, 2009 - 2018.
Member, 2001 - 2008.
- Technical Institution of Greece (TEE), 1997 - now.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

LANGUAGES

English, Greek, French.

PRIOR TESTIMONY

January 19, 2022 Deposition in *CA, Inc. et al v. Netflix, Inc.*, Case No. 2:21-cv-00080-JRG-RSP (E.D. Tex.).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

B. Table Of Opinions

§	Opinion	Pages	¶¶	Sub-Headings	Pages	¶¶
III.A.	Opinion 1: Mr. Hochman's Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect.	20-33	35-68	1. The Data At Issue Is Not Associated With A Google Account	20-25	36-47
				2. The Data At Issue Is Stored In An Orphaned And Unidentified State	25-31	48-64
				3. Privacy-Preserving Technical Barriers And Policies Prevent Google From Re-Identifying Logs Data	31-33	65-68
III.B.	Opinion 2: Mr. Hochman's Opinions On Interception, Notice, And Deletion Of Private Browsing Information (# 4, 5, 6, 26, 31) Are Contrary To Industry Guidelines On Private Browsing.	33-37	69-76			
III.C.	Opinion 3: Mr. Hochman's Opinions On "Private Browsing Profiles," Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate.	37-42	77-85	1. Orphaned And Unidentified Interest Segments Are Not Cradle-To-Grave Profiles	37-41	78-83
				2. Mr. Hochman's Claims Regarding Google's Retention Policies Show That These Policies Prevent The Creation Of Such Profiles	41-42	84-85

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	Opinion	Pages	¶¶	Sub-Headings	Pages	¶¶
III.D.	Opinion 4: Mr. Hochman's Assertion That Google Used Private Browsing Information To Measure Conversions (# 14) Is Misleading.	42-43	86-89			
III.E.	Opinion 5: Mr. Hochman's Description Of Entropy (# 18) Is Incorrect.	44-46	90-96			
III.F.	Opinion 6: Mr. Hochman's Assertions On Fingerprinting (# 9, 18) Are Misleading And Unfounded.	46-53	97-108	1. Google Does Not Engage In Fingerprinting	47-49	97-99
				2. Google's Internal Policies Expressly Prohibit Fingerprinting	49-50	100-102
				3. Google's System Architecture Supports Google's Anti-Fingerprinting Policy	51-53	103-108
III.G.	Opinion 7: Mr. Hochman's Proposal to Identify Class I (Chrome Class) (# 22) Is Unreasonable and Unreliable.	53-70	109-141	1. Hochman's IP + UA Fingerprinting Method Will Not Work	55-64	111-127
				2. Pseudonymous IDs Can Not Be Used To Reliably Identify Individuals	64-70	128-141
III.H	Opinion 8: Mr. Hochman's Opinion That The "maybe_Chrome_incognito" Bit Reliably Detects Incognito Traffic (# 23) Is Incorrect.	71-74	142-145			

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	Opinion	Pages	¶¶	Sub-Headings	Pages	¶¶
III.I	Opinion 9: Mr. Hochman’s Proposal To Identify Class II (# 22) Is Unreasonable And Unreliable.	74-83	146-161	1. Mr. Hochman’s Proposed Email Notification To All Google Account Holders Is Overly Broad And He Does Not Propose A Workable Methodology For Limiting The Notification To Class Members	76-78	148-152
				2. Mr. Hochman’s Proposed Methodology For Limiting Class II To Private Browsing Mode Users After Notification Is Unreliable	79-83	153-161
III.J	Opinion 10: Mr. Hochman’s Proposed Methods For Identifying Class Members (# 22) Do Not–And Cannot–Account For Shared Devices or Accounts.	83-95	162-180			
IV.A.	Opinion 11: Mr. Schneier’s Assertion That “Browsing Information Is Unique For Each User” Is Unsupported And Misleading.	96-99	181-186			
V.B	Opinion 12: Mr. Schneier’s Claim That “Personal Data Is Difficult to Anonymize and Easy to De-anonymize” Is Unsupported And Is Incorrect For The Data At Issue.	99-105	187-193			

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	Opinion	Pages	¶¶	Sub-Headings	Pages	¶¶
V.C.	Opinion 13: Mr. Schneier's Assertion That Google Has Not Taken Steps To Ensure That A User's Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User's Signed-Out Activity With Any Signed-In Data Is Incorrect.	105-119	194-219	1. The Documents Mr. Schneier Cites Do Not Support His Conclusion	106-109	196-202
				2. Additional Documents And Testimony Contradict Mr. Schneier's Claim That Google Has Not Undertaken Steps To Prevent Joining of Signed-Out And Signed-In Data	109-112	203-205
				3. Google's Policy Restrictions And Pseudonymization Procedures Closely Align With Best Practices For Research Involving User Data	112-115	206-211
				4. Google's Security Practices Closely Align With Best Practices In The Network Security Industry	115-119	212-218

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

C. Hochman To Psounis Mapping

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
VIII. A	Google Interception: Throughout the class period, Google intentionally intercepted private browsing communications between users and non-Google websites while those communications were in transit	20-48	A.1 Google designed tracking and advertising code to run on non-Google websites for Google's interception	82-94	# 1	"[T]hroughout the class period, Google, by way of various tracking beacons, intercepted private browsing communications between users and non-Google websites while those communications were in transit."	
			A.2 Google intercepted (and continues to intercept) private browsing communications between users and non-Google websites, with Google obtaining personal information from those communications	95-109	# 2	"[A] major function of the tracking beacons is to collect highly personal information about users' browsing activities, including users who choose a private browsing mode, such as the contents of their communications with non-Google websites in the form of detailed URL requests, webpage and video interactions, and more."	
			A.3 Google designed its tracking beacons to collect information by intercepting communications while they were in transit	110-112	# 3	"[T]hroughout the class period, the Google tracking beacons which cause private browsing communications to be intercepted neither facilitate nor are incidental to the communications between users and non-Google websites. The Google tracking beacons functioned (and continue to function) to listen in and collect information from these private browsing communications."	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
			A.4 Google's tracking beacons neither facilitate nor are they incidental to communications between users and non-Google websites	113-116			
			A.5 This functionality was not an accident; Google designed its tracking and advertising code this way	117-119	# 4	"Google could have at any point before or during the class period, redesigned Chrome Incognito to either stop or limit Google's collection of private browsing information from the private communications between users and non-Google websites."	§ III.B
			A.6 Google could have designed its products differently	120-133			
VIII. B	User Notification & Choice: Throughout class period, Google collected this private browsing information without notifying users or offering a choice at the time of or in connection with any of the interceptions	48-49	n.a.	134-135	# 5	"Google, throughout the class period, intercepted private browsing communications without notifying users or providing a choice at the time of collection. Google could have provided such notification but did not."	§ III.B

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
VIII. C	Website Notification & Choice: Throughout class period, Google collected this private browsing information without notifying websites at the time of or in connection with any of the interceptions or providing websites with a choice	49-50	n.a.	136-137	# 6	"Google, throughout the class period, intercepted private browsing communications without notifying websites or providing a choice at the time of collection. Google could have done that but did not."	§ III.B
VIII. D	Google Storage: Throughout the class period, Google stored private browsing information in many Google data sources, sometimes permanently	50-61	n.a.		# 7	"Google, throughout the class period, has stored the private browsing information it has collected in many Google data sources, sometimes permanently."	
			D.1 With the Special Master process, Google identified some but not all sources that include private browsing information	140-152	# 8	"Google, within the Special Master process, did not identify all logs or data sources that contain private browsing information, omitting key logs."	
			D.2 Google keeps some private browsing information permanently, with no option for users to review or request deletion of that information	154-161	# 9	"Google, throughout the class period, has not provided an option for users to review or request deletion of their private browsing information. Google only provided such controls for activities when users are signed into Google, while giving private browsing users no controls when they are not signed into Google."	§ III.F

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
			D.3 Google merged private and non-private browsing information in its logs and other data sources	162-165	n.a.	n.a.	§ III.B
VIII. E	Google Use: Throughout the class period, Google exploited private browsing information to generate revenue for Google by creating profiles, serving ads, tracking conversions, and in other ways that benefited Google	61-93	E.1 Google Tracking & Profiling: Throughout the class period, Google tracked private browsing communications to create detailed profiles	167-178	# 10	"Google, throughout the class period, created detailed profiles tied to various Google identifiers (that remain undisclosed to users) based on the private browsing information it collected."	§ III.C
					# 11	"Google, throughout the class period, collected valuable data (including private browsing information collected from users' visits to non-Google websites while signed out of any Google account) to feed and develop Google's machine learning algorithms, which Google in turn uses to generate advertising revenues."	
			E.2 Advertising: Throughout the class period, Google used private browsing information to serve ads (including targeted ads)	179-191	# 12	"Google's tracking beacons have, throughout the class period, enabled Google to serve advertisements to users visiting non-Google websites within their private browsing sessions while not signed into any Google account."	
					# 13	"Google, throughout the class period, used private browsing information to serve personalized advertisements to private browsing users, including on non-Google websites and while not signed into any Google account, using their private browsing	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
						information."	
			E.3 Conversion Tracking: Throughout the Class Period, Google used private browsing information to track conversions	192-211	# 14	"Google, throughout the class period, used private browsing information to measure and model conversions."	§ III.D
			E.4 Google has attempted to circumvent the blocking of Google tracking beacons by Firefox's private browsing mode and blocking of cookies by Apple's Intelligent Tracking Prevention (ITP)	212-214	# 15	"Google before and during the class period attempted to circumvent efforts by other companies to block Google tracking beacons (including with Firefox private browsing's Tracking Protection and Enhanced Tracking Protection) and to block tracking cookies (including with Apple's Intelligent Tracking Prevention)."	
			E.5 Analytics: Throughout the class period, Google used private browsing information to provide information to Google's Analytics customers and increase Google's total ads revenues	215-220	# 16	"Google, throughout the class period, used private browsing information collected through Google Analytics tracking beacons for purposes of delivering advertisements, including personalized advertisements and remarketing."	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
			E.6 Google also used private browsing information to enhance Search revenue and develop and improve other Google products and algorithms	221-222	# 17	"Google, throughout the class period, included private browsing information in the data sets used to enhance Google Search revenues and to develop and improve other Google products and algorithms."	
VIII. F	Google Joinability: Throughout the class period, Google collected and stored private browsing information in ways that can be joined to other Google user information, but Google withheld and destroyed data that would be relevant to further assessing and demonstrating that joinability	93-113	n.a.	223-258	# 18	"[T]hroughout the class period, information tied to a user's Google account could be linked to the same individual's private browsing information stored within Google logs and data sources."	§ III.A; § III.C; § III.E.; § III.F.;
					# 19	"[T]hroughout the class period, information tied to a user's account with non-Google websites could be linked to the same individual's private browsing information stored within Google logs and data sources."	§ III.C
					# 20	"Google's deletion of certain data (including data Google deleted after this lawsuit was filed) hinders the process of linking (or joining) a user's private browsing information to that user's Google account, and in some cases may make it impossible to do so where it would have been possible but for Google's deletion of the data."	§ III.C

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
VIII. G	California: Google designed its systems to provide its California-based employees with access to private browsing information collected nationwide, and Google employees routinely access that information in California	114-115	n.a.	259-261	# 21	"[T]hroughout the class period, Google's systems have provided Google's California-based employees with access to private browsing information that Google collected from users nationwide, with Google employees routinely accessing that information in California."	
VIII. H	Class Member Identification: Throughout the class period, Google collected private browsing information in ways that can be used to identify class members, though Google also withheld and destroyed data relevant to that identification	115-135	H.3 Class Identification	285-286	# 22	"[T]here are several ways to use Google's records to identify class members. For example, for Class I, the Chrome class, Google's records can be used to (1) identify Incognito traffic, and (2) link that Incognito traffic to users' Google accounts or to users' accounts with non-Google websites. As another example, for both Classes, if users were to come forward and self-identify, Google's records can be used to verify that the individual is a Google account holder who visited a non Google website that contains Google tracking beacons."	§ III.G; § III.I; § III.J
			H.3.1 Class I (Chrome Class)	287-294			
			H.3.2 All Class Members	295-307			
					# 23	"Google's 'maybe_chrome_incognito' bit reliably identifies Incognito traffic."	§ III.H
					# 24	"Google's deletion of certain data throughout the class period will hinder the class identification methods I propose, and in some cases may make it impossible to identify certain class members."	§ III.B

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
					# 25	"Google obstructed the parties' discovery efforts within the Special Master process, including by failing to identify logs that contain private browsing information, producing incomplete log schemas that omitted key fields, not properly formatting search queries and parameters, and delaying productions."	
VIII. I	Attempt: Google's attempted interception and collection uniformly impacted all class members	135-139			# 26	"Google, throughout the class period, uniformly attempted to intercept all private browsing communications with non-Google websites that have a Google tracking beacon—regardless of which private browsing mode the user employed."	§ III.B
					# 27	"Google's tracking beacons were so ubiquitous throughout the class period that there is a near certainty that almost every person using the private browsing modes at issue (in Chrome, Safari, and Edge/IE) during the class period had their private browsing information intercepted by Google, including while visiting non-Google websites without being signed into any Google account."	
VIII. J	Incognito Functionality: Throughout the class period, Google's Chrome Incognito mode functioned in ways that were different than represented	139-149	J.1 Incognito Splashscreen	319-324	# 28	"Google does not offer users any control to escape Google's tracking beacons, which are almost impossible to avoid for Internet users in the United States."	

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

§	TOC	pages	sub-heading	¶¶	#	Executive Summary	Psounis
			J.2 Google's Chrome Incognito mode functioned in ways that were different than how Google represented it	325-334	# 29	"[T]hroughout the class period, Google's Chrome Incognito mode functioned in ways that differed from how Google represented to users that Chrome Incognito functions."	
VIII. K	Incognito Changes: Google could have changed Chrome Incognito mode to match user expectations and address misconceptions, but Google chose not to	149-152			# 30	"Google, before or at any point during the class period, could have changed Chrome Incognito mode to address what Google referred to as user misconceptions, but Google did not."	
VIII. L	Google can change its processes going forward and purge its systems of private browsing information	153-155			# 31	"Google could delete from its systems records of Chrome Incognito browsing communications. Google could also delete from its systems records of users' signed-out browsing communications."	§ III.B

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

D. Psounis To Hochman And Schneier Mapping

§	Psounis Opinion	Hochman, Schneier ¶¶
III.A	Opinion 1: Mr. Hochman’s Opinion That Users Can Readily Be Identified From The Data At Issue (# 18) Is Incorrect.	<i>Hochman</i> ¶¶ 223-258
III.B	Opinion 2: Mr. Hochman’s Opinions On Interception, Notice, And Deletion Of Private Browsing Information (# 4, 5, 6, 26, 31) Are Contrary To Industry Guidelines On Private Browsing.	<i>Hochman</i> ¶¶ 78-137, 308-317, 342-349
III.C	Opinion 3: Mr. Hochman’s Opinions On “Private Browsing Profiles,” Server-Side Processes, And Data Joinability (# 10, 18, 19, 20) Are Inaccurate.	<i>Hochman</i> ¶¶ 166-258
III.D	Opinion 4: Mr. Hochman’s Assertion That Google Used Private Browsing Information To Measure Conversions (# 14) Is Misleading.	<i>Hochman</i> ¶¶ 166-222
III.E	Opinion 5: Mr. Hochman’s Description Of Entropy (# 18) Is Incorrect.	<i>Hochman</i> ¶¶ 231-233
III.F	Opinion 6: Mr. Hochman’s Assertions On Fingerprinting (# 9, 18) Are Misleading And Unfounded.	<i>Hochman</i> ¶¶ 223-258
III.G	Opinion 7: Mr. Hochman’s Proposal To Identify Class I (Chrome Class) (# 22) Is Unreasonable and Unreliable.	<i>Hochman</i> ¶¶ 262-307
III.H	Opinion 8: Mr. Hochman’s Opinion That The “maybe_Chrome_incognito” Bit Reliably Detects Incognito Traffic (# 23) Is Incorrect.	<i>Hochman</i> ¶¶ 287-294
III.I	Opinion 9: Mr. Hochman’s Proposal To Identify Class II (# 22) Is Unreasonable And Unreliable.	<i>Hochman</i> ¶¶ 262-307
III.J	Opinion 10: Mr. Hochman’s Proposed Methods For Identifying Class Members (# 22) Do Not–And Cannot–Account For Shared Devices or Accounts.	<i>Hochman</i> ¶¶ 285-307
IV.A	Opinion 11: Mr. Schneier’s Assertion That “Browsing Information Is Unique For Each User” Is Unsupported And Misleading.	<i>Schneier</i> ¶¶ 97-99
IV.B	Opinion 12: Mr. Schneier’s Claim That “Personal Data Is Difficult to Anonymize and Easy to De-anonymize” Is Unsupported And Is Incorrect For The Data At Issue.	<i>Schneier</i> ¶¶ 143-155
IV.C	Opinion 13: Mr. Schneier’s Assertion That Google Has Not Taken Steps To Ensure That A User’s Choice To Sign Out Of A Google Account Will Prevent Google From Associating The User’s Signed-Out Activity With Any Signed-In Data Is Incorrect.	<i>Schneier</i> ¶ 205

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

EXPERT REPORT OF KOSTANTINOS PSOUNIS

June 7, 2022

APPENDIX E

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

E. Entropy, Entropy Bits and Fingerprinting: Formal Exposition**1. Preliminaries: Random Experiment, Events, Probability Of Events, Random Variables, Distribution Of A Random Variable**

1. To understand how to correctly compute entropy bits of identifiers and what is their fingerprinting effectiveness, we need to start with the concept of an experiment whose outcome is random. When the experiment takes place, the outcome can be any among a set of possible events each of which is assigned a probability representing its likelihood of occurring.³¹⁷

2. The simplest possible example to consider is a coin toss. The random experiment in this example is the act of tossing a coin. There are two possible outcomes, heads or tails, and the set of possible events is thus $\{H, T\}$ where H represents heads and T represents tails. If the coin is fair, the likelihood of either H or T is the same. If we toss this coin a very large number of times and record how many times it lands heads and how many times it lands tails, we will record that 50 percent of the times we get H and 50 percent of the times we get T .³¹⁸ We thus assign a probability of 0.5 to the event H , and we write $P(H) = 0.5$, with $P(\cdot)$ representing the probability of an event, and we assign a probability 0.5 to the event T , and we write $P(T) = 0.5$.

³¹⁷ Patrick Billingsley, "Probability and Measure," Wiley, (3rd ed. 1995); Sheldon Ross, "Introduction to Probability Models," Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014).

³¹⁸ In theory, this statement is precise if we toss the coin (independently) an infinite amount of times.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

3. The next concept we need to introduce is that of a random variable.³¹⁹ A random variable is a function which maps events to probabilities. In the simple fair coin toss example, let X be the random variable representing the outcome of the coin toss. Then, the random variable takes two possible values, H or T , the probability of the events $\{X = H\}$ and $\{X = T\}$ is equal to 0.5, and we write $P(X = H) = P(X = T) = 0.5$. The coin may not be fair, for example, it may land heads way more often than tails, say 98 percent of the time. In this case, $P(X = H) = 0.98$ and $P(X = T) = 0.02$.

4. In order to introduce the concept of entropy and thus entropy bits, we need one more concept, that of the distribution of a random variable. Thankfully, for the (discrete) random variables we are considering, the distribution of a random variable is simply the probability that the random variable assumes any of its possible values. Specifically, if the random variable X may assume some values x with some probability, we define by $p_X(x)$ the distribution of the random variable X as follows:³²⁰

$p_X(x) = P(X = x)$. (Equation (1) – probability distribution)

$p_X(x)$ is a function which maps each possible value x of X to the probability that the random variable X assumes this value x . In our fair coin example, $p_X(H) = p_X(T) = 0.5$.

³¹⁹ We will restrict the discussion to discrete random variables, see Sheldon Ross, “Introduction to Probability Models,” Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014), as this is the type of random variables which are relevant to our discussion about digital fingerprinting.

³²⁰ For the discrete random variables we are considering, this distribution is formally referred to as the probability mass function, see Sheldon Ross, “Introduction to Probability Models,” Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

2. Entropy And Entropy Bits Of A Random Experiment

5. We are now ready to define the concept of entropy. The entropy $H(X)$ of a random variable X with distribution $p_X(x)$ is defined as follows:³²¹

$$H(X) = - \sum_{all\ x} p_X(x) \cdot \log_2(p_X(x)). \quad (\text{Equation (2) - entropy})$$

6. The entropy is a measure of uncertainty. The more random an experiment is, the more the uncertainty of its outcome, the larger the entropy. To see this, consider the simple coin toss example. If the coin is fair, $p_X(H) = p_X(T) = 0.5$ and Equation (2) yields $H(X) = 1$. If, however, the coin is not fair but instead $p_X(H) = 0.98$ and $p_X(T) = 0.02$, then $H(X) = 0.14$ which is a lot less than 1. As a matter of fact, if the coin is so biased that it only lands heads, i.e. $p_X(H) = 1$, then we get $H(X) = 0$. This makes sense as there is no uncertainty anymore about the outcome of the coin toss.

7. Consider again the fair coin toss example, where we calculated that the entropy $H(X) = 1$. We say that the entropy of a fair coin toss experiment is 1 entropy bit. Similarly, we say that the entropy of a coin toss with $p_X(H) = 0.98$ is 0.14 entropy bits.

8. Another useful interpretation of the entropy is that it is the expectation of $\log_2(p_X(x))$. Specifically, it is calculated as the average information conveyed by an outcome, $\log_2(p_X(x))$, averaged over all possible outcomes, x , by weighting the information conveyed by an outcome with the likelihood of this outcome, $p_X(x)$, see Equation (2).

³²¹ T.M. Cover and J.A. Thomas, "Elements of Information Theory," Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

3. Entropy Bits And Fingerprinting

9. Fingerprinting refers to the process of using an identifier to identify a member of a set. For example, an IP+UA pair may be used to identify a device which accesses a web site. In this example the identifier is the combination of an IP address and a User Agent string, and the set consists of all possible devices.

10. The first step in this process is to recognize that guessing the device that accesses the web site is the outcome of a random experiment. In the absence of any information about the devices, any device is equally likely to be the one that accesses the web site. Hence, if there are 2^K possible devices, each accessing the web site with probability $1/2^K$, using Equation (2) yields that the entropy equals K entropy bits.

11. The second step in this process is to recognize that observing a particular identifier value is the outcome of another random experiment. To compute the entropy of the identifier, one needs to know, or estimate from a dataset, the distribution of the values of the identifier, and then use Equation (2). In general, we want the entropy bits of the identifier to be as large as possible, hopefully as large as K . That said, having many entropy bits is only a necessary condition, it is not sufficient, as we also need to have a way to go from identifier values to devices.

12. The third step in this process is to have a mapping which reliably maps the observed value of the identifier to a single device. If there is no reliable mapping relating an identifier value to a device, because, for example, an identifier value may correspond to multiple devices, then, regardless of the entropy bits of the identifier, obviously we cannot reliably identify a device, see [§ V.E.7](#) below for a simple example..

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

a. A simple numerical example

13. Let's consider a simple example to illustrate some of the points we have discussed. Suppose that our identifier is a UA, there are only two possible UA values, say UA_1 and UA_2 , and we are trying to guess which web browser among, say, $64 = 2^6$ possible web browsers, is the one with the UA we just observed. In the absence of any information, any of the 64 web browsers are equally likely to be the one, each with probability $1/64$. Let X denote the random variable representing the web browser, say $X=1,2,\dots,64$. We have $p_X(x) = 1/64$ for all x and, using Equation (2), $H(X) = 6$, that is, our uncertainty is six entropy bits.

14. The fingerprinting effectiveness of an identifier has to do with how much it can reduce this uncertainty. If $P(UA = UA_1) = P(UA = UA_2) = 0.5$, the identifier has 1 entropy bit and can reduce the uncertainty from 6 bits to $6-1=5$ entropy bits. This makes intuitive sense: Since $P(UA_1) = P(UA_2) = 0.5$, it must be that 32 of the 64 web browsers have UA_1 , and 32 have UA_2 . Thus, knowing that $UA = UA_1$ restricts the set of possible browsers from 64 to 32, and the remaining entropy is 5 as the remaining $32 = 2^5$ web browsers are equally likely to be the one.

15. How much would the UA identifier decrease the uncertainty if its two possible values are not equally likely but instead one is much more likely than the other? As an example, assume that $P(UA_1) = 63/64$, because 63 of the 64 web browsers have this UA value. We already know the answer to this, as $63/64$ is about 0.98 and we have already computed the entropy of a binary random variable with distribution $\{0.98, 0.02\}$ to be equal to 0.14.³²² Hence,

³²² $63/64=0.984375 > 0.98$, so, as a matter of fact, the identifier is even weaker than our calculations here, though by very little.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

the uncertainty is now reduced from 6 bits to $6 - 0.14 = 5.86$ entropy bits. Intuitively, the reason why the identifier is so ineffective now is because with probability 0.98 we will observe UA_1 and this will give us very little information, we will merely restrict the set of possible browsers from 64 to 63. While it is true that if we observe a UA_2 we will uniquely identify one of the 64 web browsers, the probability of this event is so low that, on average, we are really learning very little from this identifier.

16. Notice that even though the identifier only has 0.14 entropy bits, if we observe UA_2 we will uniquely identify one web browser. This has to do with the fact that entropy is an expectation thus the entropy of an identifier represents its average ability to perform fingerprinting. An identifier with low entropy may perform bad on average but it does not guarantee that no browser can be reliably identified. Similarly, an identifier with high entropy may perform well on average but it does not guarantee that all browsers can be reliably identified.

17. The fact that more skewed distributions yield less entropy bits is not merely a numerical observation. It can be proved that a uniform distribution, that is, a distribution where all outcomes are equally likely, maximizes uncertainty and thus entropy bits,³²³ and the more skewed a distribution is, the deterministic case being the most skewed of all as only one outcome is possible (with probability 1), the less the uncertainty and the less the entropy bits (0 in the deterministic case).

³²³ T.M. Cover and J.A. Thomas, “Elements of Information Theory,” Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

4. More Preliminaries: Independence, Conditional Probability, Jointly Distributed Random Variables

18. To understand how to compute and interpret the joint entropy of more than one identifier and what is the fingerprinting effectiveness of multiple identifiers used together, we first need to introduce the concepts of independence, conditional probability, and jointly distributed random variables.

19. Two events are independent if the outcome of one does not affect the outcome of the other. As an example, if one has two coins in one's pocket, tosses the first coin, and then tosses the second coin, it makes intuitive sense to argue that whether the first coin lands heads or tails (first event) does not affect whether the second coin lands heads or tails (second event). If we denote by H_1 the event that the first coin lands heads, and by H_2 the event that the second coin lands heads, then, formally, H_1 and H_2 are independent if:

$$P(H_1 \text{ and } H_2) = P(H_1) \cdot P(H_2), \quad (\text{Equation (3) – independent events})$$

where $P(H_1)$ is the probability of the event H_1 , $P(H_2)$ is the probability of the event H_2 , and $P(H_1 \text{ and } H_2)$ is the probability that both H_1 and H_2 occur, a.k.a. the intersection of the two events.

20. The intuition about the independence of two events carries over for random variables as well. If we use X_1 to denote the random variable associated with the first coin toss, and X_2 to denote the random variable associated with the second coin toss, one may argue that X_1 and X_2 are independent random variables, in the sense that the value that one assumes does not affect the value that the other assumes, or, knowing what value one assumes, does not change our guess about what value the other may assume.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

21. To formally define the concept of independent random variables (and later the concept of joint entropy which is directly related with the fingerprinting effectiveness of multiple identifiers used together) we first define the concept of a joint distribution of two random variables X_1 and X_2 as follows:

$$p_{X_1, X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2). \quad (\text{Equation (4) – joint distribution})$$

Similar to the case of one random variable, the joint distribution of two random variables is a function that yields the probability that $X_1 = x_1$ and $X_2 = x_2$ for all possible values x_1 and x_2 .

In our simple example of two coin tosses, there are four possible joint cases to consider, $\{H_1, H_2\}$ (both coins land heads), $\{H_1, T_2\}$, $\{T_1, H_2\}$ and $\{T_1, T_2\}$, and the joint distribution gives the probability of each of these four cases.

22. We are now ready to formally define the concept of independent random variables. Two random variables X_1 and X_2 are independent if and only if:

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2), \quad (\text{Equation (5) – independent random variables})$$

that is, the joint distribution of the two random variables equals the product of the two individual distributions of each random variable.³²⁴ If the two coin tosses are independent and both coins are fair, it is easy to see that each of the 4 possible outcomes occurs with probability $1/2 \cdot 1/2 = 1/4$.

23. Not all events / random variables are independent. For example, suppose X_1 represents whether there are clouds in the sky on a particular day ($X_1 = 1$ indicates there are

³²⁴ Individual distributions are called marginal distributions.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

clouds and 0 otherwise) and X_2 represents whether it will rain this day ($X_2 = 1$ indicates it will rain and 0 otherwise). Knowing that there are clouds in the sky does increase the chance that it will rain hence X_1 and X_2 are dependent. The concept of conditional probability allows us to figure out how this dependence changes the corresponding probabilities. Let C represent the event “clouds in the sky” and R the event “rain”. We formally define the conditional probability of event R given event C , i.e. the probability R occurs given that we know C occurred, as follows:

$$P(R|C) = P(R \text{ and } C)/P(C), \quad (\text{Equation (6) – conditional probability})$$

where $P(C)$ is the probability of the event that there are clouds, considered non-zero, and $P(R \text{ and } C)$ is the probability of the event that there are clouds and it is raining, i.e. the intersection of the two events. The key here is to observe that it is more likely to rain if there are clouds in the sky, hence $P(R|C) > P(R)$, where $P(R)$ is the unconditional probability of whether it will rain or not. Also, one may argue that it is impossible to rain without clouds, in which case one may argue that $P(R|C) = 1$, since, given that there is rain, there must be clouds.

24. Note that if two events, H_1 and H_2 , are independent, then from Equations (3) and (6) we get that $P(H_2|H_1) = P(H_2)$, which makes intuitive sense as knowledge of whether H_1 has occurred or not does not have any effect on the likelihood of H_2 , since they are independent.

25. Going back to the concept of the joint distribution, with X_1 and X_2 representing clouds and rain respectively, it is obvious that the 4 possible events {clouds, rain}, {clouds, no rain}, {no clouds, rain}, {no clouds, no rain} are not equally likely. For example, it is

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

impossible to rain without clouds hence the probability of the outcome {no clouds, rain} equals $P(X_1 = 0, X_2 = 1) = 0$.

26. Last, we define the conditional distribution of a random variable X_2 given the value of another random variable X_1 , as a function that yields the probability that X_2 may assume a value x , given that X_1 has assumed a known value x_1 . Formally, the conditional distribution of X_2 given X_1 is given by:

$$p_{X_2|X_1}(x|x_1) = p_{X_2, X_1}(x, x_1)/p_{X_1}(x_1), \quad (\text{Equation (7) – conditional distribution})$$

for all possible values x of X_2 , where $p_{X_1}(x_1)$ is non-zero. Note that if X_1 and X_2 are independent, Equations (5) and (7) imply that $p_{X_2|X_1}(x|x_1) = p_{X_2}(x)$, that is, the conditional distribution equals the individual (marginal) distribution. This makes intuitive sense, as conditioning on the value of X_1 does not change the distribution of X_2 since they are independent.

5. Joint Entropy And Conditional Entropy

27. We have defined entropy in Equation (2) and explained how entropy bits are directly related to the effectiveness of an identifier to perform fingerprinting by reducing uncertainty. To understand how to calculate the entropy and entropy bits of multiple identifiers used together, we first define the concepts of joint entropy and conditional entropy.³²⁵

28. The joint entropy of two random variables X_1 and X_2 is defined as follows:

³²⁵ T.M. Cover and J.A. Thomas, “Elements of Information Theory,” Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

$$H(X_1, X_2) = - \sum_{all\ x_1, x_2} p_{X_1, X_2}(x_1, x_2) \cdot \log_2(p_{X_1, X_2}(x_1, x_2)). \quad (\text{Equation (8) – joint entropy})$$

The joint entropy, similar to the entropy of a single random variable, measures the uncertainty of a pair of random variables. Note that because $P(X_1 = x_1, X_2 = x_2) = P(X_2 = x_2, X_1 = x_1)$, it follows that $p_{X_1, X_2}(x_1, x_2) = p_{X_2, X_1}(x_2, x_1)$ and $H(X_1, X_2) = H(X_2, X_1)$.

29. The conditional entropy of X_2 given X_1 is defined as follows:

$$H(X_2|X_1) = - \sum_{all\ x_1, x_2} p_{X_1, X_2}(x_1, x_2) \cdot \log_2(p_{X_2|X_1}(x_2|x_1)). \quad (\text{Equation (9) – conditional entropy})$$

To understand the intuition behind conditional entropy, note the following equation which relates the joint entropy of two random variables X_1 and X_2 with the entropy of X_1 alone:³²⁶

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1). \quad (\text{Equation (10) – meaning of conditional entropy})$$

Equation (10) implies that $H(X_2|X_1)$ is the additional uncertainty in $\{X_1, X_2\}$ due to X_2 , when the uncertainty of X_1 is already accounted for.

30. It can be shown that $H(X_2|X_1) \leq H(X_2)$ with equality if and only if they are independent. Thus, from Equation (10) we conclude that $H(X_1, X_2) = H(X_1) + H(X_2)$ if and only if X_1 and X_2 are independent.

31. There are two methods to compute $H(X_1, X_2)$. One is to use Equation (8) if the joint distribution of X_1 and X_2 is known, or it can be estimated from a dataset. The other is to

³²⁶ *Id.*

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

use Equation (10), which is particularly easy to do if X_1 and X_2 are independent, since in this case the joint entropy is simply the sum of the individual entropies.

6. Entropy Bits And Fingerprinting Effectiveness Of Multiple Identifiers

32. We can now explain the fingerprinting effectiveness of multiple identifiers used together. Without loss of generality, consider two identifiers, represented by the random variables X_1 and X_2 , since we do not know in advance which value the identifiers may assume. First, let's consider the case where the identifiers are independent. In this case, $H(X_1, X_2) = H(X_1) + H(X_2)$ and the total entropy bits of the pair of identifiers equals the sum of the individual entropy bits, where, as already discussed, the individual entropy bits depend on how skewed the distributions of X_1 and X_2 are.

33. Now, let's consider the case where the identifiers are not independent. In this case, $H(X_1, X_2) = H(X_1) + H(X_2|X_1)$. The first term in this equation depends on how skewed the distribution of X_1 is. The second term depends on how skewed the conditional distribution of X_2 is, given X_1 . Even if the marginal distribution of X_2 is, say, uniform, which, as discussed, would maximize the entropy $H(X_2)$, the conditional distribution may be very skewed. As a matter of fact, conditioning on X_1 may even reduce the possible values that X_2 may assume to 1, yielding a deterministic case and zero additional entropy bits.

a. A numerical example involving multiple identifiers

34. We present a simple numerical example involving two popular identifiers often used jointly, the UA and the IP address. Say X_1 represents the UA and X_2 the IP address of a device / web browser. Suppose there are a total of 64 possible browsers, thus the entropy bits

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

associated with guessing which browser accesses a web site is 6 bits, since $2^6 = 64$ and all browsers, in the absence of any other information, are equally likely to be the one. How much can we reduce the uncertainty of this estimate by observing the UA and the IP address? This depends not only on how skewed the individual distributions of the two identifiers are, p_{X_1} and p_{X_2} , but also on how dependent they are, which is captured by the conditional distribution $p_{X_2|X_1}$.

35. Like in our previous example, assume there are two possible UA values, that is, X_1 is equal to UA_1 or UA_2 . Also, assume that there are $4 = 2^2$ possible IP addresses, that is, X_2 is equal to IP_1, IP_2, IP_3, IP_4 . If X_1 and X_2 have uniform distributions and are independent, then the entropy bits of the two identifiers combined is $1+2 = 3$, and the uncertainty of guessing the web browser is reduced from 6 to $6-3 = 3$ bits. We have essentially narrowed down our guess from 64 to 8 possible browsers. But, in practice, none of these assumptions may hold: the distributions of both X_1 and X_2 are skewed,³²⁷ and X_1 and X_2 may not be independent.³²⁸

36. To illustrate the drastic effect of lifting the uniform assumption, assume that the distribution of X_1 equals, like before, $\{0.98, 0.02\}$, that is, $P(X_1 = UA_1) = 0.98$ and $P(X_1 = UA_2) = 0.02$. Then, $H(X_1) = 0.14$, which is a lot less than 1 entropy bit (uniform distribution case). Also, assume the distribution of X_2 is skewed, say equal to $\{0.48, 0.48, 0.02, 0.02\}$, because, for example, 31 out of the 64 web browsers have the same IP address equal to

³²⁷ J. Kline, P. Barford, A. Cahn, and G. Sommers, "On the structure and characteristics of user agent strings", ACM Sigcomm, <https://conferences.sigcomm.org/imc/2017/papers/imc17-final253.pdf> (Nov. 1-3, 2017), at Figure 4.

³²⁸ For example, if devices have the same IP address because they sit behind a VPN, it is likely to also share the same UA if the devices belong to the same company and the IT department is updating the software of the company devices in batches.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

IP_1 , another 31 have the same IP address IP_2 , and the last two web browsers have IP addresses IP_3 and IP_4 respectively.³²⁹ Then, from Equation (2) we compute $H(X_2) = 1.24$, which is sizably less than 2 entropy bits (uniform distribution case). Even if we assume X_1 and X_2 are independent, to get the maximum entropy, the combined fingerprinting effectiveness of UA and IP is reduced from 3 bits (uniform case) to merely $0.14 + 1.24 = 1.38$, representing a more than 55 percent reduction in fingerprinting effectiveness.

37. To illustrate the drastic effect of lifting the independence assumption, we will first assume that X_1 is uniform, that is, UA_1 and UA_2 are equally likely, such that $H(X_1) = 1$. This is the maximum possible entropy bits for UA. And, we will also assume that X_1 and X_2 are not independent. Specifically, assume that all 31 browsers with IP address equal to IP_1 have the same UA,³³⁰ equal to UA_1 , all 31 browsers with IP address equal to IP_2 have the same UA, equal to UA_2 , and for the remaining 2 browsers with IP addresses IP_3 and IP_4 , we assume it is equally likely to have a UA equal to either UA_1 or UA_2 . Essentially, we described the conditional distribution of X_2 given X_1 to be $\{31/64, 0, 1/64, 1/64\}$ when $X_1 = UA_1$, and $\{0, 31/64, 1/64, 1/64\}$ when $X_1 = UA_2$. We can now use Equations (7) and (9) and compute $H(X_2|X_1) = 0.22$. Notice how few entropy bits X_2 adds given X_1 . This is because UA (X_1) reveals a lot about the IP (X_2), for example, $X_1 = UA_1$ informs us that the set of possible browsers is very likely to be one of the 31 with IP_1 and for sure none of the browsers with IP_2 .

³²⁹ This may be the case because 31 IP addresses connect to the internet via the same VPN server and thus share the same IP address, IP_1 , and the same is true for IP_2 .

³³⁰ As discussed in the main text, when devices share the same IP address, e.g. because they are behind a VPN server of an organization, it is likely they also share UAs, as the IT department of the organization may centrally control the software updates on organizations' devices.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Equation (1) yields $H(X_1, X_2) = 1 + 0.22 = 1.22$, which is again a lot less than the 3 bits under the assumptions of uniform distributions and independence, representing an almost 60 percent reduction in fingerprint-ability.

7. On The Entropy Bits Reported By Mr. Hochman

38. Mr. Hochman states that “[w]ith around 330 million people in the United States, 29 bits of data is more than sufficient to identify a person ($2^{29} = 537$ million).”³³¹ He states that “IP address and User Agent ‘carries 29.8 bits’ of entropy, which is more than sufficient to uniquely identify individuals in the United States.”³³²

39. Mr. Hochman’s description of entropy as a measure for user identifiability is incorrect, see *infra* [§ V.E.2](#). Entropy is a measure of the uncertainty of the outcome of a random process. Entropy can not be used to establish what specific information is necessary to uniquely identify any specific person, or how many bits of data are “more than sufficient to uniquely identify individuals in the United States” in the absence of a one to one mapping which maps identifier values to specific individuals, see *infra* [§ V.E.3](#), and the example below in [§ V.E.7.a](#).

40. Neither Hochman nor the document he cites, GOOG-CABR-04635379, provide any information about the mathematical methodology or dataset used to derive the 29.8 entropy bits value number for the IP address and User Agent pair. This is required to establish credibility for the reported number of entropy bits. At a minimum, is the dataset restricted to individuals in the United States or is it collected by Google from one of its services offered worldwide and thus this dataset relates to the global population? In an attempt to predict which individual in the U.S. (1 out of 330 million) is accessing a private browsing session, in the

³³¹ Hochman ¶ 231.

³³² Hochman ¶ 233, citing GOOG-CABR-04635379 at -380. See also Hochman ¶ 231 (asserting that “Google engineers measure that IP address alone contributes to 26.5 bits of entropy (GOOG-BRWN-00601937).”).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

absence of any information, and assuming it is equally likely for any individual in the US to be the one accessing a private browsing session, applying the definition of entropy for 330 million people yields an entropy of 28.3 bits ($2^{28.3} \approx 330M$). Doing the same exercise with a global population of 8 billion yields 32.9 bits ($2^{32.9} \approx 8B$), which is obviously larger than 29.8.

41. As discussed already in [§ V.E.3](#), there are two entropy bits calculations related to fingerprinting. The first concerns the entropy of the task at hand, e.g. identifying an individual accessing a web site among the U.S. population. The second concerns the entropy of the identifier used to do so, which depends on the number of distinct values it may assume and on how skewed the distribution of the popularity of these values is.

42. A necessary but not sufficient condition for an identifier to perform well on average is to have more entropy bits than the entropy of the task at hand. Contrary to what Mr. Hochman states, we do not know if the IP+UA pair satisfies the necessary condition, because we do not know if the dataset is collected over the global population, or over the U.S. population, or any other population. But we do know for sure that it does not satisfy the requirement for a reliable one-to-one mapping of IP+UA values to individuals. A fundamental reason for this is that an IP+UA pair may correspond to multiple users because: (i) devices belonging to different users may have the same IP+UA, see *infra* [§ III.G.1](#), and (ii) users may share a device, see *infra* [§ III.J](#).

a. A simple mapping example

43. Consider three persons, John, Mary, and Peter. John and Mary live in the same house. John owns a phone and shares a home laptop with Mary. Mary owns a phone, the same model as John's, and a work laptop. When at home, John's and Mary's devices connect to the home WiFi and sit behind a NAT. Mary and Peter are co-workers. Peter owns a phone and a

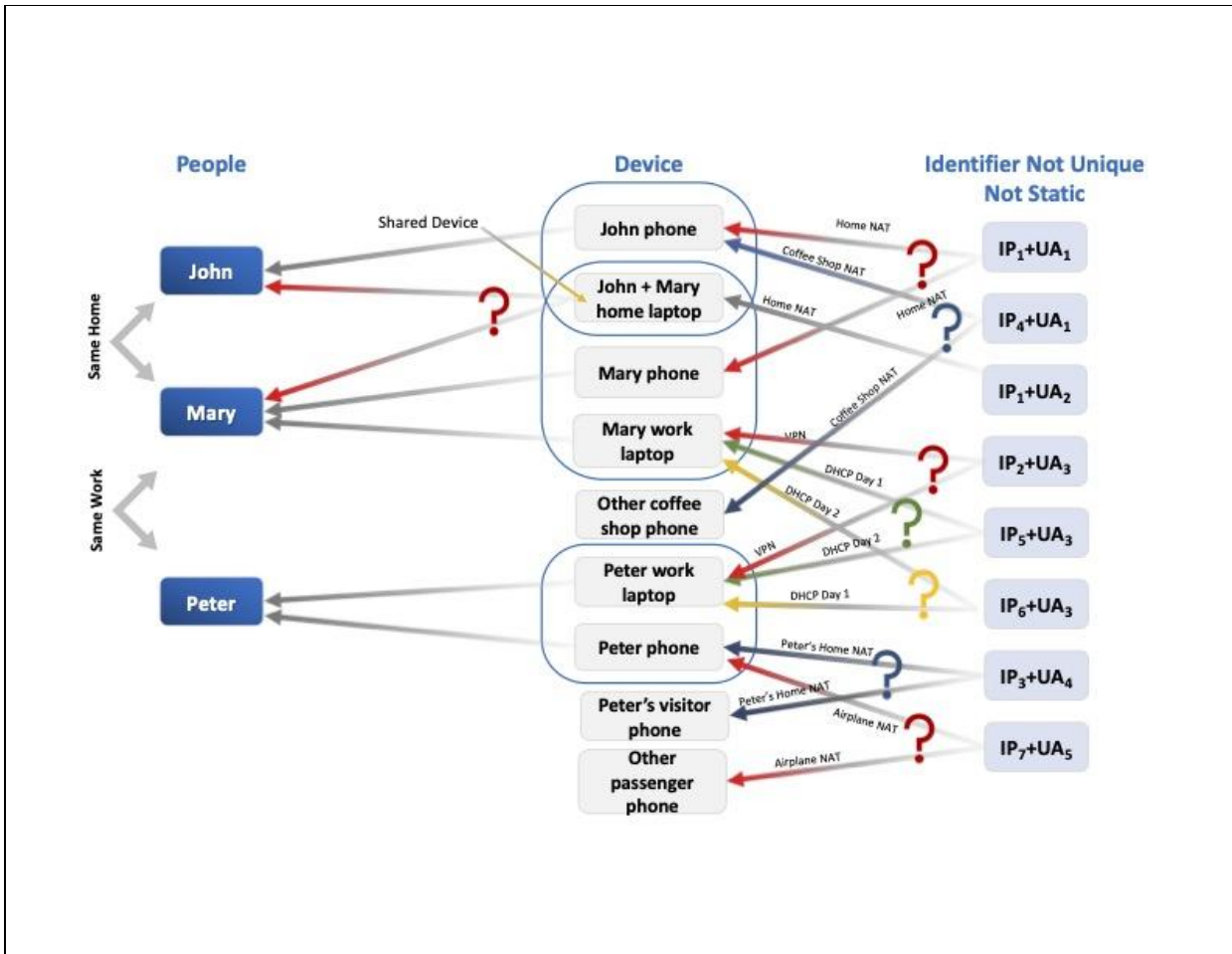
CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

work laptop. Sometimes they work from home and connect to their employer's network via the same VPN server. When they work in the office, their employer uses DHCP to assign IP addresses to their work laptops, and the IT department updates the software on the company's devices.³³³

44. The following graph shows the complex mapping between IP+UA values, devices, and persons. John's and Mary's phones, when at home, share the same IP+UA. John's and Mary's browsing sessions via their shared home laptop share the same IP+UA. John's and another coffee shop client's phone may share the same IP+UA, if the other client has the same phone model as John. Mary's and Peter's work laptops, when working from home, share the same IP+UA. Mary's and Peter's work laptops, when in the office, may have the same IP+UA on different days. Peter's phone may share the same IP+UA with the phone of a visitor in his home or a passenger on the same plane, if the other person has the same phone model as Peter. In all these cases, merely involving less than a handful of people, it is impossible to uniquely identify an individual through an IP+UA.

³³³ See [§ III.G.1](#) for more information on NAT, VPN and DHCP.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

EXPERT REPORT OF KOSTANTINOS PSOUNIS

June 7, 2022

APPENDIX F

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

F. IP Address + User Agent Data Analysis

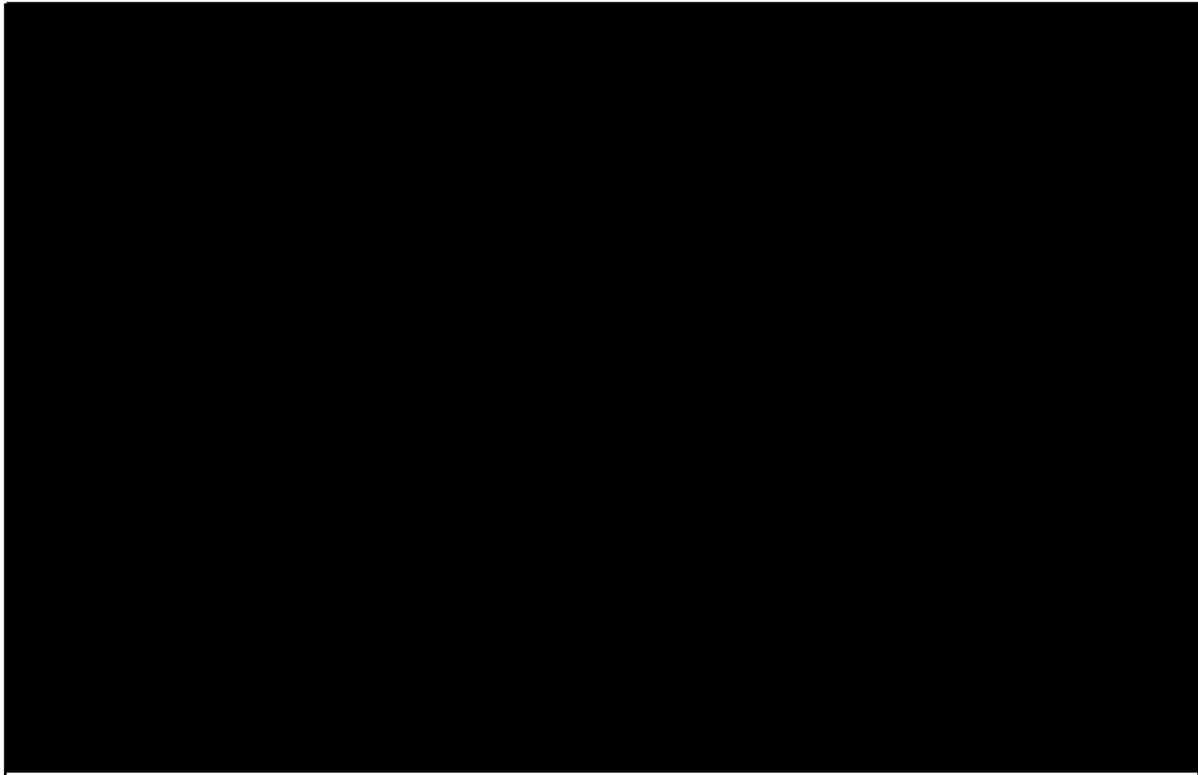
1. I understand that Google preserved authenticated data associated with seven GAIA IDs provided by the five named Plaintiffs from relevant data sources pursuant to a litigation hold and produced to Plaintiffs on March 9, 2022 the preserved data from all relevant logs that do not contain confidential third-party information (GOOG-BRWN-00847947-GOOG-BRWN-00847948).

2. The retained data (in total 144,937 records) contains **4,945** distinct IP addresses, **three** of which correspond to multiple different Plaintiffs' GAIA IDs:

IP Address	Corresponding GAIA ID/Plaintiff	Timestamp
[REDACTED]	[REDACTED]	[REDACTED]
	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]
	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]
	[REDACTED]	[REDACTED]

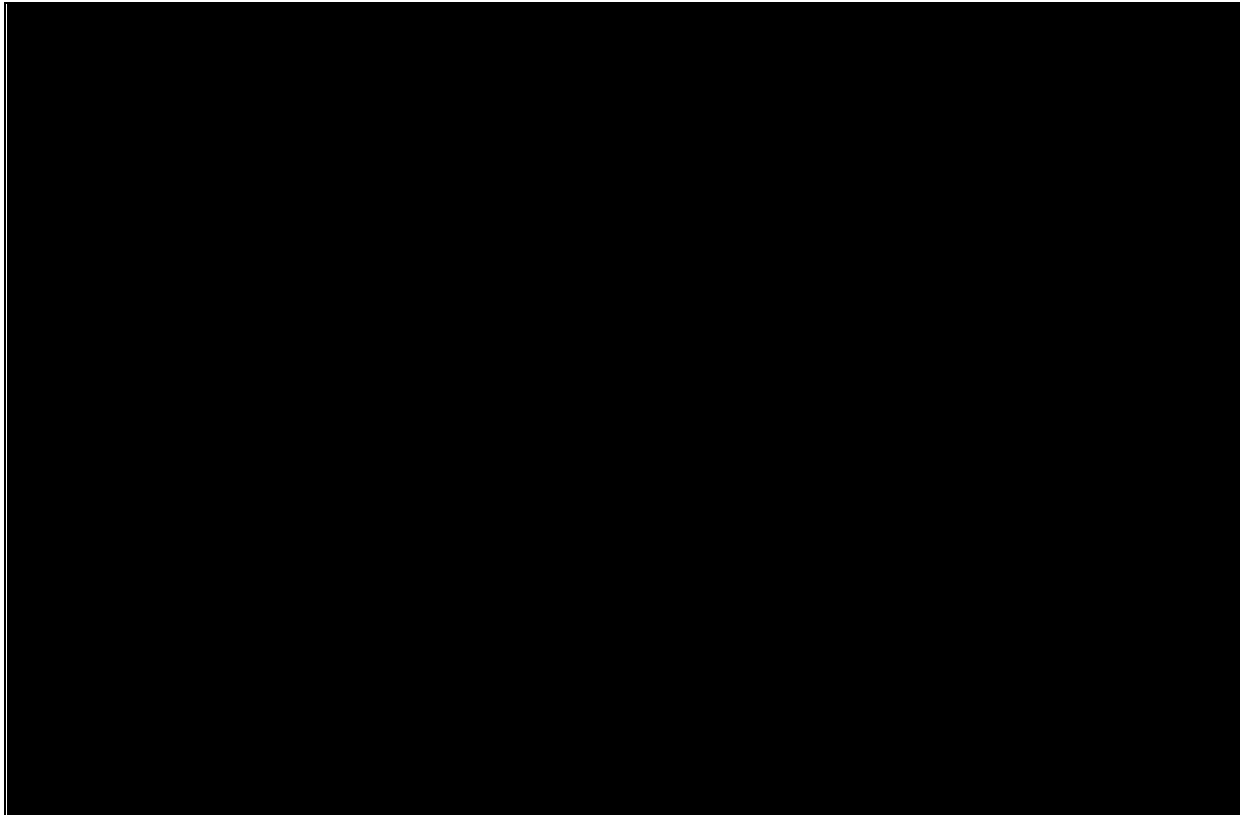
3. According to the public IP address lookup tool available at <https://whatismyipaddress.com/>, IP address [REDACTED] corresponds to ISP T-mobile USA Inc. located in Los Angeles, California:

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



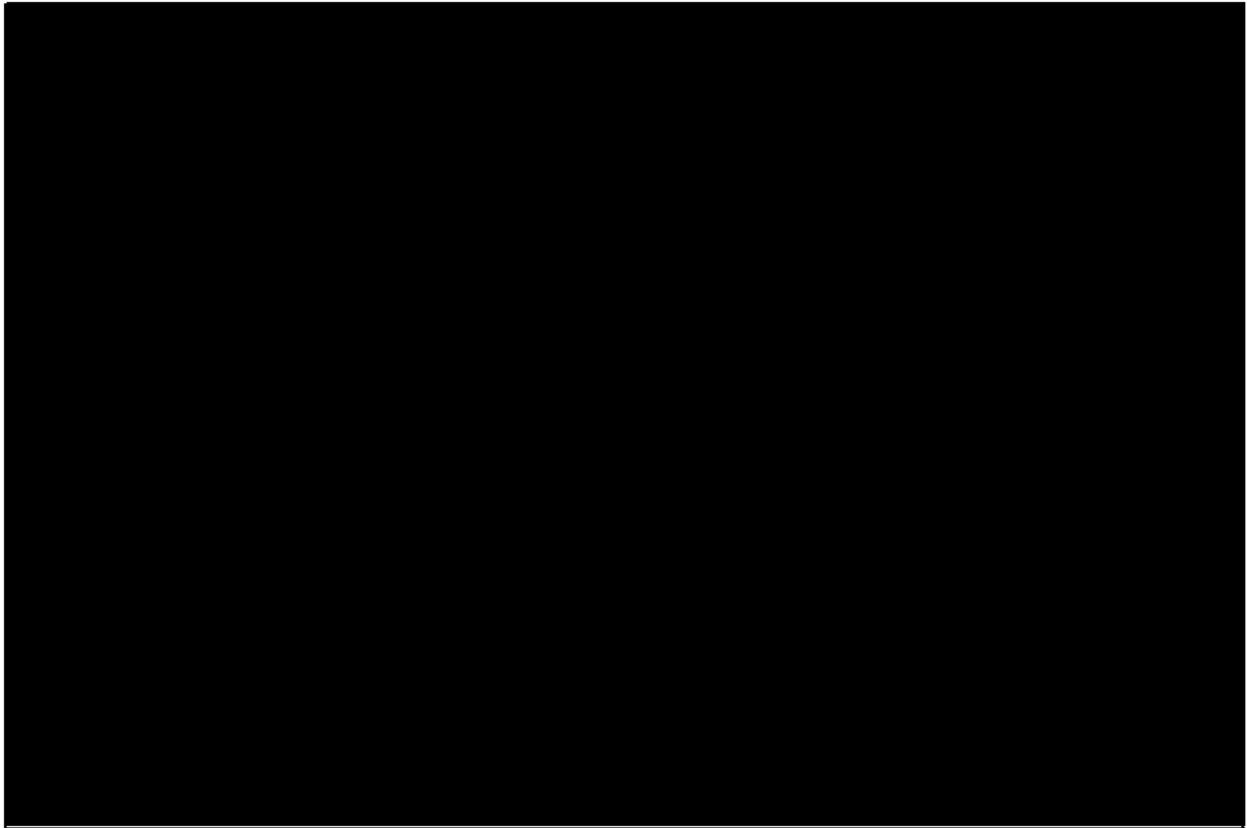
4. According to the public IP address lookup tool available at <https://whatismyipaddress.com/>, IP address [REDACTED] also corresponds to ISP T-mobile USA Inc. located in Los Angeles, California:

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



5. According to the public IP address lookup tool available at <https://whatismyipaddress.com/>, IP address [REDACTED] corresponds to ISP T-mobile USA Inc. located in Dallas, Texas:

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



6. The data also contains **1,237** distinct User Agent strings, **73** of which correspond to multiple different Plaintiffs' GAIA IDs:

User Agent	Corresponding GAIA ID & Plaintiff
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

179

180

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

User Agent	Corresponding GAIA ID & Plaintiff
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

User Agent	Corresponding GAIA ID & Plaintiff
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

User Agent	Corresponding GAIA ID & Plaintiff
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

User Agent	Corresponding GAIA ID & Plaintiff
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

7. The data also contains **8,965** unique IP address and UA combinations. For details, *see* backup material titled “IP + UA Analysis (Source Material: GOOG-BRWN-00847947-948) (CONFIDENTIAL).”

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

EXPERT REPORT OF KOSTANTINOS PSOUNIS

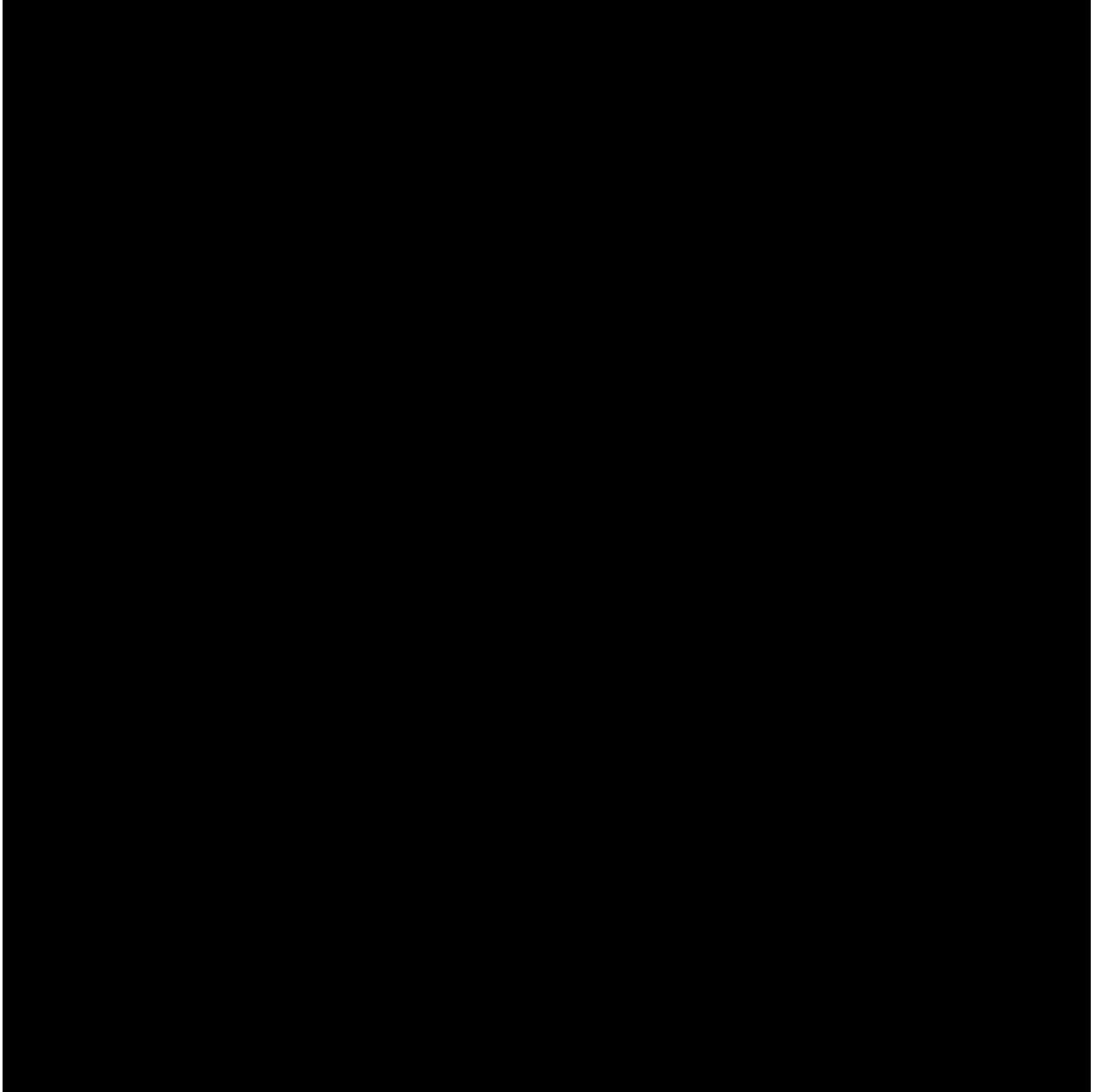
June 7, 2022

APPENDIX G

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

G. “Profile” Data

1. Below is a side-by-side comparison of the produced DBL [REDACTED] data stored in the column [REDACTED] on the same day associated with the two different Biscotti IDs³³⁴. I used Google produced files to decode the information and added my notes in red.

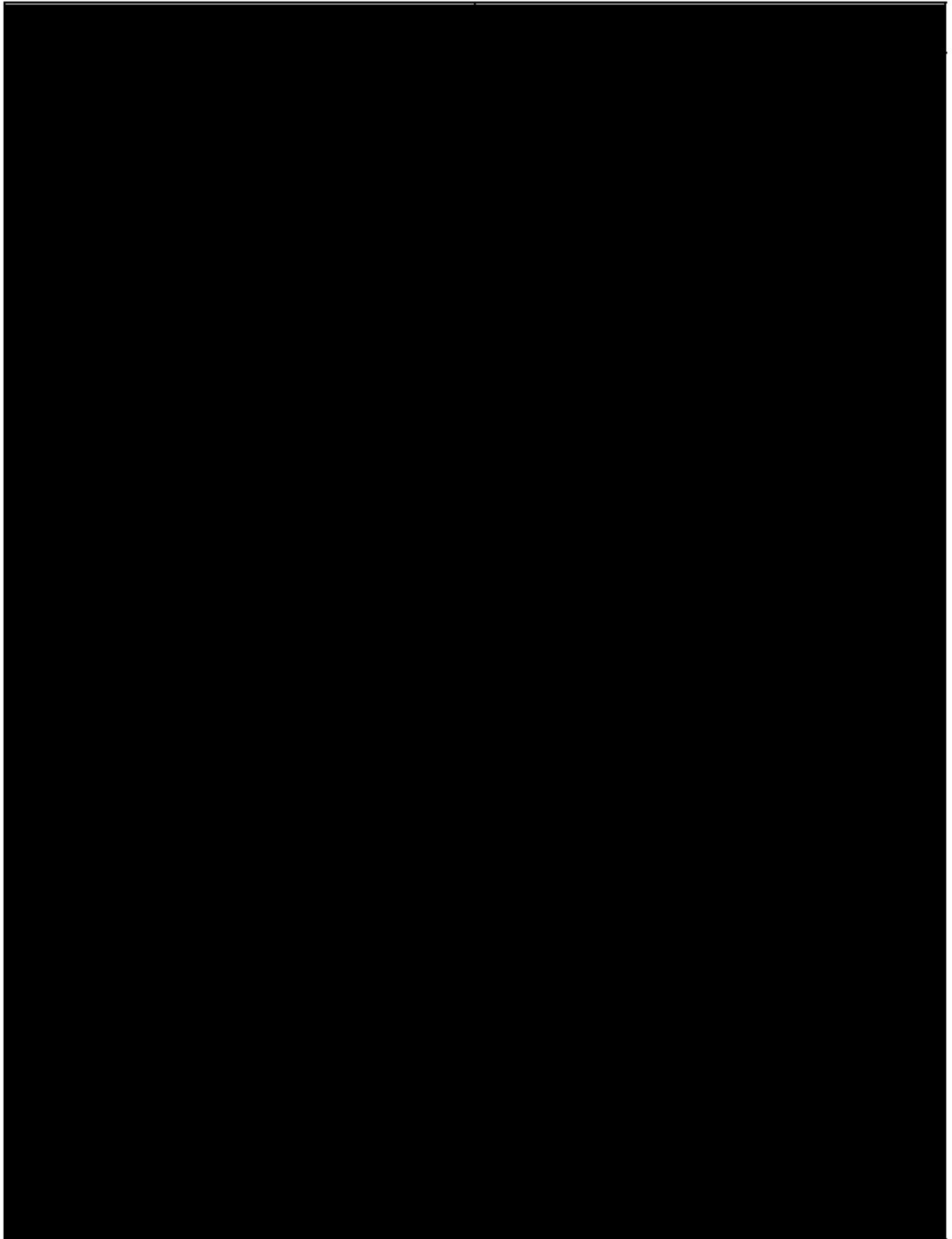


³³⁴ See GOOG-BRWN-00229628.

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

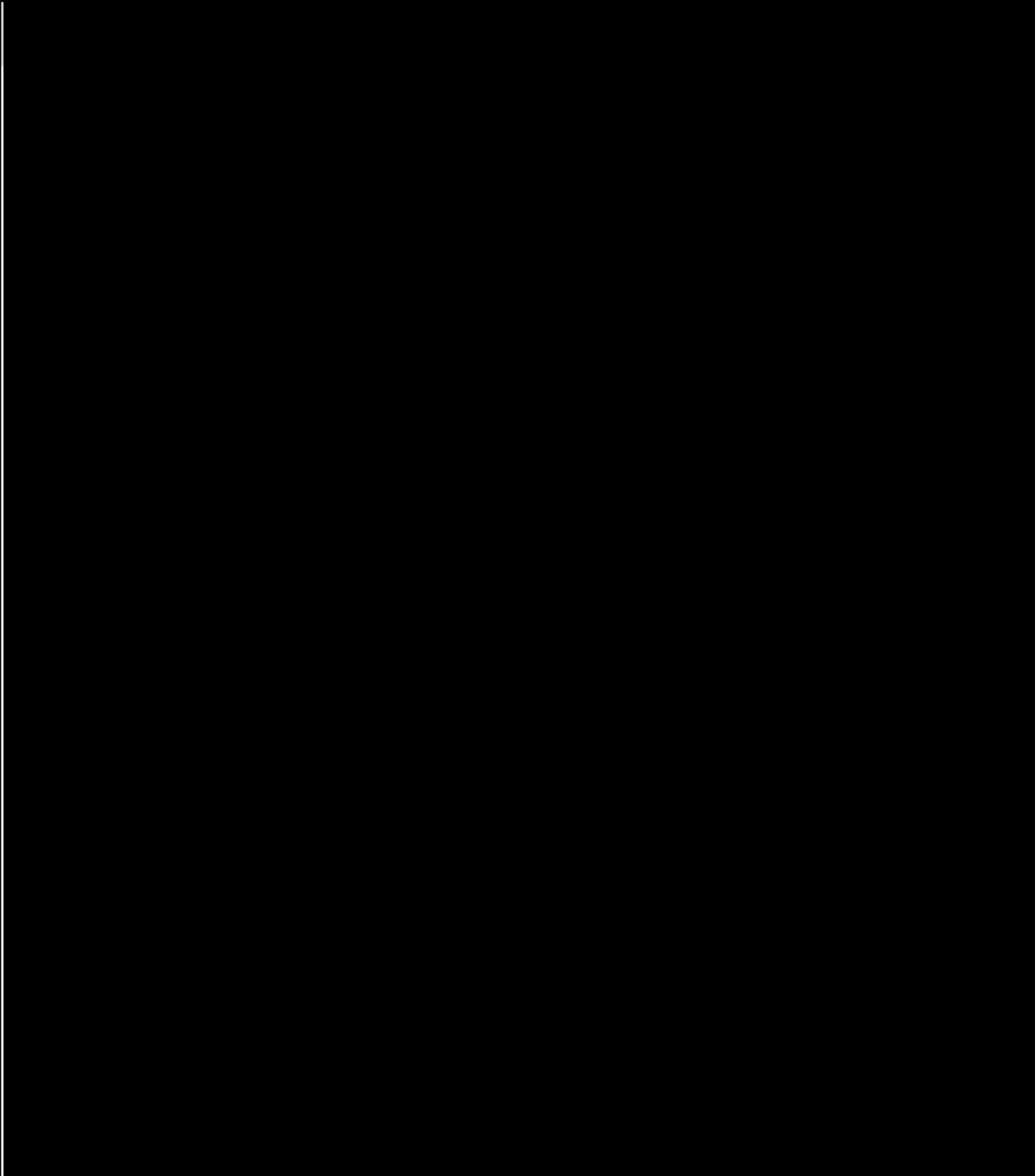


CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER



CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

2. Below is a side-by-side comparison of the decoded inferred interest segments pulled from the table above:



CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

H. Sources Considered

Filed Documents:

- Third Amended Class Action Complaint, Chasom Brown, et al., v. Google LLC, United States District Court Northern District of California, February 3, 2022 (Dkt. 395-2)
- Microsoft's Response to Brown Plaintiffs' Subpoena (Aug. 27, 2021)
- Mozilla's Response to Brown Plaintiffs' Subpoena (Aug. 27, 2021)
- Apple's Response to Brown Plaintiffs' Subpoena (Sept. 20, 2021)
- Google's Supplemental Objections and Responses to Plaintiffs' Interrogatories (No. 17) (Apr. 7, 2022)
- Declaration of Jonathan E. Hochman, Telebrands Corp., v. Tinnus Enterprises, LLC, Case No. PGR2015-00018 (USPTO 2015)
- 2022-05-09 - Plaintiff Brown Attestation
- 2022-05-09 - Plaintiff Castillo Attestation
- 2022-05-09 - Plaintiff Trujillo Attestation

Expert Reports:

- Apr. 15, 2022 Expert Report of Jonathan E. Hochman and Appendices and Exhibits Attached Thereto
- Apr. 15, 2022 Expert Report of Bruce Schneier and Appendices and Exhibits Attached Thereto
- Expert Report of Jonathan E. Hochman, *Rockwood Select Asset Fund XI, (6)-1, LLC v. Devine, Millimet & Branch, P.A.*, Case No. 1:14-cv-00303-LM, 2016 WL 4260622 (D.N.H. Jan. 29, 2016)

Depositions and Hearings:

- April 14, 2022 Special Master Hearing Transcript
- April 21, 2022 Hearing Transcript
- Deposition Transcript of Abdelkarim Mardini, Nov. 24, 2021
- Deposition Transcript of Adrienne Porter-Felt, Nov. 16, 2021

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Deposition Transcript of Alexei Svitkine, Oct. 4, 2021
- Deposition Transcript of Brian Rakowski, Aug. 19, 2021
- Deposition Transcript of Chasom Brown, Jan. 13, 2022
- Deposition Transcript of Chetna Bindra, Feb. 8, 2022
- Deposition Transcript of Chris Palmer, Jan. 5, 2022
- Deposition Transcript of Christopher Castillo, Feb. 11, 2022
- Deposition Transcript of David Monsees, Apr. 9, 2021
- Deposition Transcript of David Monsees, June 11, 2021
- Deposition Transcript of Caitlin Sadowski, Mar. 10, 2022
- Deposition Transcript of Glenn Berntson, June 16, 2021
- Deposition Transcript of Glenn Berntson, Mar. 18, 2022
- Deposition Transcript of Gregory Lon Fair, Dec. 14, 2021
- Deposition Transcript of Huei-Hung Chris Liao, Dec. 3, 2021
- Deposition Transcript of Hyewon Jun, Mar. 1, 2022
- Deposition Transcript of Jeremy Davis, Jan. 7, 2022
- Deposition Transcript of Jesse Adkins, Apr. 4, 2021
- Deposition Transcript of Justin Schuh, Jan. 6, 2022
- Deposition Transcript of Mandy Liu, Mar. 8, 2022
- Deposition Transcript of Martin Shelton, Mar. 2, 2022
- Deposition Transcript of Michael Kleber, Mar. 18, 2022
- Deposition Transcript of Michael Kleber, Jan. 14, 2022
- Deposition Transcript of Monique Trujillo, Feb. 22, 2022
- Deposition Transcript of Ramin Halavati, Jan. 18, 2022
- Deposition Transcript of Rory McClelland, Feb. 18, 2022
- Deposition Transcript of Sammit Adhya, Nov. 19, 2021
- Deposition Transcript of Stephen Chung, Mar. 10, 2022

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Deposition Transcript of Steve Ganem, Mar. 23, 2022
- Deposition Transcript of William Byatt, Dec. 20, 2021
- Deposition Transcript of Wing Pan “Bert” Leung, Mar. 4, 2022

Public Documents:

- M. Clark and K. Psounis, “Optimizing Primary User Privacy in Spectrum Sharing Systems,” *IEEE/ACM Transactions on Networking*, <https://ieeexplore.ieee.org/document/8985324> (Apr. 2020)
- J. Zhang, K. Psounis, M. Zaroon, and Z. Shafiq, “HARPO: Learning to Subvert Online Behavioral Advertising,” *NDSS*, <https://web.cs.ucdavis.edu/~zubair/files/harpo-ndss2022.pdf> (Apr. 2022)
- T. Spyropoulos, K. Psounis, and C. Raghavendra, “Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-copy Case,” *IEEE/ACM Transactions on Networking*, <https://cpb-us-e1.wpmucdn.com/sites.usc.edu/dist/b/364/files/2019/05/multiton.pdf> (Feb. 2008)
- T. Spyropoulos, K. Psounis, and C. Raghavendra, “Efficient Routing in Intermittently Connected Mobile Networks: The Single-Copy Case,” *IEEE/ACM Transactions on Networking*, <https://ee.usc.edu/netpd/assets/001/51984.pdf> (Feb. 2008)
- Alex Sherman, “Netflix estimates 100 million households are sharing passwords and suggests a global crackdown is coming,” *CNBC*, <https://www.cnbc.com/2022/04/19/netflix-warns-password-sharing-crackdown-is-coming.html> (Apr. 20, 2022)
- Matt Richtel, “Young, in Love and Sharing Everything, Including a Password,” *N.Y. Times*, <https://www.nytimes.com/2012/01/18/us/teenagers-sharing-passwords-as-show-of-affection.html> (Jan. 17, 2012)
- Arjun Ruparelia, “Best DraftKings Sportsbook VPN in 2022: Unblock DraftKings From Anywhere With a VPN,” *Cloudwards*, <https://www.cloudwards.net/draftkings-sportsbook-vpn/> (Apr. 27, 2022)
- Osman Husain, “How to watch ESPN anywhere with a VPN,” *Comparitech*, <https://www.comparitech.com/blog/vpn-privacy/best-vpn-espn/> (Jan. 19, 2022)
- M. Levin and J. Lowitz, “iPhone 13 Models Have Biggest Share in Years,” *Consumer Intelligence Research Partners LLC*,

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<https://files.constantcontact.com/150f9af2201/06eda1e6-cb00-4462-836f-73aa0120e439.pdf?rdr=true> (Apr. 21, 2022)

- S. O’Dea, “App-updating habits of smartphone users in the United States 2016,” Statista, <https://www.statista.com/statistics/747569/united-states-survey-smartphone-users-app-update-frequency/#statisticContainer> (Feb. 28, 2020)
- J. Kline, P. Barford, A. Cahn, and G. Sommers, “On the structure and characteristics of user agent strings,” ACM Sigcomm, <https://conferences.sigcomm.org/imc/2017/papers/imc17-final253.pdf> (Nov. 1-3, 2017).
- E. Rye, R. Beverly, and K. Claffy, “Follow the Scent: Defeating IPv6 Prefix Rotation Privacy,” Proceedings of ACM Internet Measurement Conference (IMC), <https://arxiv.org/pdf/2102.00542.pdf> (Nov. 2-4, 2021).
- D. Plonka and A. Berger, “Temporal and Spatial Classification of Active IPv6 Addresses,” Proceedings of ACM Internet Measurement Conference (IMC), <https://www.akamai.com/site/en/documents/research-paper/temporal-and-spatial-classification-of-active-ipv6-addresses-technical-publication.pdf> (Oct. 28-30, 2015).
- Tara Matthews, et. al., “‘She’ll just grab any device that’s closer’: A Study of Everyday Device and Account Sharing in Households,” Proceedings of the ACM Conference on Human Factors in Computing Systems, ACM, <https://dl.acm.org/doi/10.1145/2858036.2858051> (2016)
- K. Levy and B. Schneier, “Privacy threats in intimate relationships,” 6 Journal of Cybersecurity 1, <https://academic.oup.com/cybersecurity/article/6/1/tyaa006/5849222> (2020)
- A. Brush and K. Inkpen, “Yours, Mine and Ours? Sharing and Use of Technology in Domestic Environments,” Proceedings of the 9th International Conference on Ubiquitous Computing, <https://www.microsoft.com/en-us/research/wp-content/uploads/2007/09/brushinkpenyoursmineours.pdf> (2007)
- B. Busse and M. Fuchs, “Prevalence of Cell Phone Sharing,” Survey Methods: Insights from the Field, <https://surveyinsights.org/?p=1019> (2013)
- H. Muller, J. Gove, and J. Webb, “Understanding Tablet Use: A Multi-Method Exploration,” Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services, <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38135.pdf> (2012).
- Rick Paulus, “The Digital Divide Is About Much More Than Access,” Pacific Standard, <https://psmag.com/environment/digital-divide-more-complicated-than-access> (June 14, 2017)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- C. Park, et al., “Share and share alike? An exploration of secure behaviors in romantic relationships,” Fourteenth Symposium on Usable Privacy and Security (SOUPS), https://www.researchgate.net/publication/325608530_Share_and_Share_Alike_An_Exploration_of_Secure_Behaviors_in_Romantic_Relationships (2018).
- Juliette Garside, “Ofcom: six-year-olds understand digital technology better than adults,” The Guardian, <https://www.theguardian.com/technology/2014/aug/07/ofcom-children-digital-technology-better-than-adults> (Aug. 6, 2014)
- Laveh Waddel, “Will Today’s Kids Be Stumped by the Technology of the Future?” The Atlantic, <https://www.theatlantic.com/technology/archive/2016/01/will-todays-kids-be-stumped-by-the-technology-of-the-future/425082/> (Jan. 26, 2016)
- Research Compliance Office, Stanford University, HRPP Policy Manual.pdf, <https://researchcompliance.stanford.edu/panels/hs/policies> (last visited June 3, 2022).
- Kamala D. Harris, “California Data Breach Report,” Cal. Dept. of Justice, <https://oag.ca.gov/breachreport2016> (2016),
- M. Fischer, J. Hochman, and D. Boffa, “Privacy-Preserving Data Sharing for Medical Research,” International Symposium on Stabilization, Safety, and Security of Distributed Systems, <https://cpsc.yale.edu/sites/default/files/files/TR1558.pdf> (November 17–20, 2021)
- L. Olejnik, C. Castellucia, and A. Janc, “Why Johnny can’t browse in peace: On the uniqueness of web browsing history patterns,” Annals of Telecommunications 1-2, <https://hal.inria.fr/file/index/docid/747841/filename/johnny2hotpet-finalcam.pdf> (June 2013)
- A. Janc and L. Olejnik, “Web browser history detection as a real-world privacy threat,” ESORICS’10: Proceedings of the 15th European Conference on Research in Computer Security, <http://cds.cern.ch/record/1293097/files/LHCb-PROC-2010-036.pdf> (Sept. 20, 2010)
- S. Englehardt, et al., “Cookies that give you away: The surveillance implications of web tracking,” WWW ‘15: Proceedings of the 24th International Conference on World Wide Web, https://senglehardt.com/papers/www15_cookie_surveil.pdf (May 18, 2015)
- Natasha Lomas, “France fines Google \$120M and Amazon 42M for dropping tracking cookies without consent,” Tech Crunch, <https://techcrunch.com/2020/12/10/france-fines-google-120m-and-amazon-42m-for-dropping-tracking-cookies-without-consent> (Dec. 10, 2020)
- Paul Ohm, “Broken promises of privacy: Responding to the surprising failure of anonymization,” UCLA Law Review 57, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 (August 13, 2009).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Latanya Sweeney, “Simple demographics often identify people uniquely,” Carnegie Mellon University Data Privacy Working Paper 3, <https://dataprivacylab.org/projects/identifiability/paper1.pdf> (2000)
- Philippe Golle, “Revisiting the uniqueness of simple demographics in the U.S. population,” 5th ACM Workshop on Privacy in the Electronic Society (WPES’06), Alexandria, Virginia, <https://crypto.stanford.edu/~pgolle/papers/census.pdf> (Oct. 30, 2006)
- L. Sweeney, A. Abu and J. Winn, “Identifying participants in the Personal Genome Project by name (A re-identification experiment),” arxiv.org, <https://arxiv.org/abs/1304.7605> (2013)
- Latanya Sweeney, “Only you, your doctor, and many others may know,” Technology Science 2018, <https://techscience.org/a/2015092903> (Sep. 28, 2015)
- Ji Su Yoo, et al., “Risks to patient privacy: A re-identification of patients in Maine and Vermont statewide hospital data,” Technology Science 2018, <https://techscience.org/a/2018100901> (Oct. 8, 2018)
- Katherine E. Boronow, et al., “Privacy risks of sharing data from environmental health studies,” Environmental Health Perspectives 128, no. 1, <https://ehp.niehs.nih.gov/doi/10.1289/EHP4817> (January 2020)
- M. Barbaro and T. Zeller Jr., “A face is exposed for AOL Search No. 4417749,” New York Times, <http://www.nytimes.com/2006/08/09/technology/09aol.html> (August 9, 2006)
- A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” 2008 IEEE Symposium on Security and Privacy, Oakland, California, <https://web.stanford.edu/class/cs245/win2020/readings/netflix-deanonymization.pdf> (May 18-20, 2008)
- Yves-Alexandre de Montjoye, et al., “Unique in the shopping mall: On the re-identifiability of credit card metadata,” Science 347, no. 6221, <https://www.science.org/doi/full/10.1126/science.1256297> (Jan. 30, 2015)
- L. Rocher, J. Jendrickx and Y. de Montjoye, “Estimating the success of re-identifications in incomplete datasets using generative models,” Nature Communications 10, <https://www.nature.com/articles/s41467-019-10933-3> (July 23, 2019).
- Douglas C. Schmidt, et al., “Google data collection,” Vanderbilt University, <https://digitalcontentnext.org/wpcontent/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf> (Aug. 15, 2018)
- Dániel Kondor, et al., “Towards matching user mobility traces in large-scale datasets,” arXiv:1709.05772, <https://arxiv.org/pdf/1709.05772.pdf> (Aug. 13, 2018)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Eli Jacobson, et al., “De-identification is insufficient to protect student privacy, or What can a field trip reveal?” *Journal of Learning Analytics* 8, no 2, <https://www.learning-analytics.info/index.php/JLA/article/view/7353> (2021)
- Pern Hui Chia, et al., “KHyperLogLog: Estimating reidentifiability and joinability of large data at scale,” *Proceedings of the IEEE Symposium on Security and Privacy*, <https://milinda-perera.com/pdf/CDPSLDWG19.pdf> (2019)
- P. Laperdrix, W. Rudametkin, and B. Baudry, “Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints,” *IEEE Symposium on Security and Privacy*, <https://ieeexplore.ieee.org/abstract/document/7546540> (2016)
- Y. Cao, S. Li, and E. Wijmans, “(Cross-)Browser Fingerprinting via OS and Hardware Level Features,” *NDSS*, San Diego CA, https://www.ndss-symposium.org/wp-content/uploads/2017/09/ndss2017_02B-3_Cao_paper.pdf (2017)
- Patrick Billingsley, “Probability and Measure,” Wiley, (3rd ed. 1995); Sheldon Ross, “Introduction to Probability Models,” Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014)
- Sheldon Ross, “Introduction to Probability Models,” Academic Press, https://www.academia.edu/17872355/Introduction_to_Probability_Models_Tenth_Edition (10th ed. 2014)
- T.M. Cover and J.A. Thomas, “Elements of Information Theory,” Wiley, https://www.academia.edu/25024538/Elements_of_Information_Theory_2nd_ed_T_Cover_J_Thomas_Wiley_2006_WW (2nd ed. 2006)
- T. Narten, R. Draves, and S. Krishnan, “Privacy Extensions for Stateless Address Autoconfiguration in IPv6,” RFC 4941 (Draft Standard), <https://www.rfc-editor.org/rfc/pdf/rfc4941.txt.pdf> (Sept. 2007)
- Tomek Mrugalski, et al., “Dynamic Host Configuration Protocol for IPv6 (DHCPv6),” RFC 8415 (Proposed Standard), <https://www.rfc-editor.org/rfc/pdf/rfc8415.txt.pdf> (Nov. 2018).
- J.F. Kurose & K.W. Ross, *Computer Networking: A Top-Down Approach* Ch. 4 (8th ed. 2020).
- W3C, <https://www.w3.org> (last visited June 3, 2022).
- W3C, “W3C Tag,” <https://tag.w3.org> (last visited June 3, 2022).
- W3C, “W3C TAG Observations on Private Browsing Modes,” W3C TAG Finding, <https://www.w3.org/2001/tag/doc/private-browsing-modes/> (July 5, 2019)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Chromium Code Search, “user_action_event.proto,”
[http://xn--https-hw3b//source.chromium.org/chromium/chromium/src/+main:third_party/metrics_proto/user_action_event.proto;l=21-27?q=time_sec%20\(last%20visited%20June%203,%202022\)](http://xn--https-hw3b//source.chromium.org/chromium/chromium/src/+main:third_party/metrics_proto/user_action_event.proto;l=21-27?q=time_sec%20(last%20visited%20June%203,%202022)).
- Chromium Code Search, “time.h,”
<https://source.chromium.org/chromium/chromium/src/+main:base/time/time.h;l=16-22?q=TimeTicks> (last visited June 3, 2022).
- NordVPN, <https://nordvpn.com/features/dedicated-ip/> (last visited June 3, 2022)
- Microsoft, "Dynamic Host Configuration Protocol (DHCP)," <https://perma.cc/7N5L-QKCQ> (July 29, 2021)
- MDN Plus, "User-Agent," <https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/User-Agent> (last visited June 2, 2022)
- Chrome Developers, "User-Agent Strings," <https://developer.chrome.com/docs/multidevice/user-agent/> (updated Nov. 9, 2021)
- Microsoft, “High Availability,”
<https://docs.microsoft.com/en-us/windows-server/remote/remote-access/vpn/always-on-vpn/deploy/always-on-vpn-adv-options#high-availability> (last visited June 3, 2022).
- Aleksandar Kochovski, “The Top 25 VPN Statistics, Facts & Trends for 2022,”
<https://www.cloudwards.net/vpn-statistics/> (Mar. 18, 2022).
- Low Entropy, “Entropy and Privacy Analysis,”
<https://lowentropy.net/posts/entropy-privacy/> (last visited June 3, 2022).
- GoogleIPv6, <https://www.google.com/intl/en/ipv6/statistics.html> (last visited June 3, 2022).
- Nord VPN, <https://nordvpn.com/features/hide-ip/> (last visited June 3, 2022)
- Surfshark, <https://surfshark.com/use-cases> (last visited June 3, 2022)
- Day One Services, “Maintaining Privacy When You Browse The Internet,”
<http://dayoneservices.org/be-safe/> (last visited June 3, 2022)
- Sahara Cares, “Protect Yourself Online,” <https://saharacares.org/protect-yourself-online/> (last visited June 3, 2022)
- WNY Postpartum Connection, Inc., “Domestic Abuse Support,”
<https://www.wnypostpartum.com/domestic-violence-assistance> (last visited June 3, 2022)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- Betty Griffin Center, “Safe Browsing,” <https://bettygriffincenter.org/help/> (last visited June 3, 2022)
- Center for Internet Security, “CIS Critical Security Controls Version 8” <https://www.cisecurity.org/controls/v8> (last visited June 3, 2022)
- National Science Foundation, “Human Subjects,” <https://www.nsf.gov/bfa/dias/policy/human.jsp> (last visited June 3, 2022)
- National Institutes of Health, “Protecting Sensitive Data and Information Used in Research,” https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.12_protecting_sensitive_data_and_information_used_in_research.htm (updated Dec. 2021)
- Office for the Protection of Research Subjects, USC, <https://oprs.usc.edu/policies/> (last visited June 3, 2022)
- Office for the Protection of Research Subjects, USC, “Chapter 10: Privacy, Confidentiality and HIPAA,” <https://oprs.usc.edu/policies/privacy-confidentiality-and-hipaa/> (last visited June 3, 2022).
- Center for Internet Security, “CIS Critical Security Controls FAQ,” <https://www.cisecurity.org/controls/cis-controls-faq> (last visited June 3, 2022)
- Google Ad Manager Help, “About publisher provided identifiers,” <https://support.google.com/admanager/answer/2880055?hl=en>, (last visited June 6, 2022)
- Google Analytics Help, “About User-ID views,” <https://support.google.com/analytics/answer/3123669?hl=en> (last visited June 6, 2022)
- Merriam-Websters.com Dictionary, ““From (the) cradle to (the) grave.”” <https://www.merriam-webster.com/dictionary/from%20%28the%29%20cradle%20to%20%28the%29%20grave> (last visited June 6, 2022).
- Epoch Converter, “Epoch & Unix Timestamp Conversion Tools” <https://www.epochconverter.com/> (last visited June 6, 2022).
- Twitter, <https://twitter.com/Jehochman/status/1153277584542711808> (last visited June 3, 2022)
- Lawrence Abrams, “How to Switch Back to the Old Twitter Layout,” Bleeping Computer, <https://www.bleepingcomputer.com/news/technology/how-to-switch-back-to-the-old-twitter-layout/> (July 16, 2019))
- W3C, <https://www.w3.org> (last visited June 3, 2022).

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- W3C, “W3C Tag,” <https://tag.w3.org> (last visited June 3, 2022).
- W3C, “W3C TAG Observations on Private Browsing Modes,” W3C TAG Finding, <https://www.w3.org/2001/tag/doc/private-browsing-modes/> (July 5, 2019)
- Chromium Code Search, “user_action_event.proto,” [http://xn--https-hw3b//source.chromium.org/chromium/chromium/src/+/main:third_party/metrics_proto/user_action_event.proto;l=21-27?q=time_sec%20\(last%20visited%20June%203,%202022\)](http://xn--https-hw3b//source.chromium.org/chromium/chromium/src/+/main:third_party/metrics_proto/user_action_event.proto;l=21-27?q=time_sec%20(last%20visited%20June%203,%202022)).
- Chromium Code Search, “time.h,” <https://source.chromium.org/chromium/chromium/src/+/main:base/time/time.h;l=16-22?q=TimeTicks> (last visited June 3, 2022).
- NordVPN, <https://nordvpn.com/features/dedicated-ip/> (last visited June 3, 2022)
- Microsoft, “Dynamic Host Configuration Protocol (DHCP),” <https://perma.cc/7N5L-QKCQ> (July 29, 2021)
- MDN Plus, “User-Agent,” <https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/User-Agent> (last visited June 2, 2022)
- Chrome Developers, “User-Agent Strings,” <https://developer.chrome.com/docs/multidevice/user-agent/> (updated Nov. 9, 2021)
- Microsoft, “High Availability,” <https://docs.microsoft.com/en-us/windows-server/remote/remote-access/vpn/always-on-vpn/deploy/always-on-vpn-adv-options#high-availability> (last visited June 3, 2022).
- Aleksandar Kochovski, “The Top 25 VPN Statistics, Facts & Trends for 2022,” <https://www.cloudwards.net/vpn-statistics/> (Mar. 18, 2022).
- Low Entropy, “Entropy and Privacy Analysis,” <https://lowentropy.net/posts/entropy-privacy/> (last visited June 3, 2022).
- GoogleIPv6, <https://www.google.com/intl/en/ipv6/statistics.html> (last visited June 3, 2022).
- Nord VPN, <https://nordvpn.com/features/hide-ip/> (last visited June 3, 2022)
- Surfshark, <https://surfshark.com/use-cases> (last visited June 3, 2022)
- Day One Services, “Maintaining Privacy When You Browse The Internet,” <http://dayoneservices.org/be-safe/> (last visited June 3, 2022)
- Sahara Cares, “Protect Yourself Online,” <https://saharacares.org/protect-yourself-online/> (last visited June 3, 2022)

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

- WNY Postpartum Connection, Inc., “Domestic Abuse Support,”
<https://www.wnypostpartum.com/domestic-violence-assistance> (last visited June 3, 2022)
- Betty Griffin Center, “Safe Browsing,” <https://bettygriffincenter.org/help/> (last visited June 3, 2022)
- Center for Internet Security, “CIS Critical Security Controls Version 8”
<https://www.cisecurity.org/controls/v8> (last visited June 3, 2022)
- National Science Foundation, “Human Subjects,”
<https://www.nsf.gov/bfa/dias/policy/human.jsp> (last visited June 3, 2022)
- National Institutes of Health, “Protecting Sensitive Data and Information Used in Research,”
https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.12_protecting_sensitive_data_and_information_used_in_research.htm (updated Dec. 2021)
- Office for the Protection of Research Subjects, USC, <https://oprs.usc.edu/policies/> (last visited June 3, 2022)
- Office for the Protection of Research Subjects, USC, “Chapter 10: Privacy, Confidentiality and HIPAA,”
<https://oprs.usc.edu/policies/privacy-confidentiality-and-hipaa/> (last visited June 3, 2022).
- Center for Internet Security, “CIS Critical Security Controls FAQ,”
<https://www.cisecurity.org/controls/cis-controls-faq> (last visited June 3, 2022)
- Google Ad Manager Help, “About publisher provided identifiers,”
<https://support.google.com/admanager/answer/2880055?hl=en>, (last visited June 6, 2022)
- Google Analytics Help, “About User-ID views,”
<https://support.google.com/analytics/answer/3123669?hl=en> (last visited June 6, 2022)
- Merriam-Websters.com Dictionary, ““From (the) cradle to (the) grave.””
<https://www.merriam-webster.com/dictionary/from%20%28the%29%20cradle%20to%20%28the%29%20grave> (last visited June 6, 2022).
- Epoch Converter, “Epoch & Unix Timestamp Conversion Tools”
<https://www.epochconverter.com/> (last visited June 6, 2022).
- Twitter, <https://twitter.com/Jehochman/status/1153277584542711808> (last visited June 3, 2022)
- Lawrence Abrams, “How to Switch Back to the Old Twitter Layout,” Bleeping Computer,

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

<https://www.bleepingcomputer.com/news/technology/how-to-switch-back-to-the-old-twitter-layout/> (July 16, 2019))

- U.S. Patent No. 7,603,483B2 (issued Oct. 13, 2009)
- National Science Foundation, SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet, Award# (USC): 1956435
- National Science Foundation, CNS Core: Medium: Collaborative Research: Privacy-Preserving Mobile Crowdsourcing, Award# (USC): 1901488
- National Science Foundation, NeTS: Spectrum Sharing Systems for Wireless Networks: Performance and Privacy Challenges

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

Produced Documents

GOOG-BRWN-00016645	GOOG-BRWN-00072761	GOOG-BRWN-00204684
GOOG-BRWN-00023909	GOOG-BRWN-00078348	GOOG-BRWN-00207020
GOOG-BRWN-00026812	GOOG-BRWN-00078361	GOOG-BRWN-00226894
GOOG-BRWN-00026913	GOOG-BRWN-00140297	GOOG-BRWN-00229628
GOOG-BRWN-00026989	GOOG-BRWN-00144320	GOOG-BRWN-00229632
GOOG-BRWN-00027090	GOOG-BRWN-00147864	GOOG-BRWN-00232201
GOOG-BRWN-00027102	GOOG-BRWN-00147873	GOOG-BRWN-00386402
GOOG-BRWN-00027314	GOOG-BRWN-00148029	GOOG-BRWN-00386511
GOOG-BRWN-00027368	GOOG-BRWN-00148738	GOOG-BRWN-00386570
GOOG-BRWN-00028052	GOOG-BRWN-00149515	GOOG-BRWN-00388293
GOOG-BRWN-00029002	GOOG-BRWN-00149849	GOOG-BRWN-00390418
GOOG-BRWN-00029326	GOOG-BRWN-00152199	GOOG-BRWN-00397243
GOOG-BRWN-00029378	GOOG-BRWN-00153850.C	GOOG-BRWN-00406065
GOOG-BRWN-00029433	GOOG-BRWN-00155943	GOOG-BRWN-00408322
GOOG-BRWN-00029445	GOOG-BRWN-00156752	GOOG-BRWN-00410076
GOOG-BRWN-00032906	GOOG-BRWN-00157001	GOOG-BRWN-00410821
GOOG-BRWN-00033024	GOOG-BRWN-00157528	GOOG-BRWN-00424253
GOOG-BRWN-00041778	GOOG-BRWN-00160342	GOOG-BRWN-00426550
GOOG-BRWN-00042388	GOOG-BRWN-00161432	GOOG-BRWN-00432838
GOOG-BRWN-00047390	GOOG-BRWN-00162235	GOOG-BRWN-00433264
GOOG-BRWN-00047399	GOOG-BRWN-00164056	GOOG-BRWN-00433503
GOOG-BRWN-00048773	GOOG-BRWN-00164626	GOOG-BRWN-00441285
GOOG-BRWN-00048967.C	GOOG-BRWN-00165626	GOOG-BRWN-00457255
GOOG-BRWN-00051239	GOOG-BRWN-00165706	GOOG-BRWN-00466617
GOOG-BRWN-00060463	GOOG-BRWN-00166360	GOOG-BRWN-00466897
GOOG-BRWN-00061607	GOOG-BRWN-00176433	GOOG-BRWN-00467569
GOOG-BRWN-00066643	GOOG-BRWN-00176481	GOOG-BRWN-00468530
GOOG-BRWN-00072051	GOOG-BRWN-00184875	GOOG-BRWN-00468598
GOOG-BRWN-00072055	GOOG-BRWN-00186446	GOOG-BRWN-00471401

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

GOOG-BRWN-00475063	GOOG-BRWN-00708814	GOOG-BRWN-00847297
GOOG-BRWN-00490767	GOOG-BRWN-00711561	GOOG-BRWN-00847947
GOOG-BRWN-00491711	GOOG-BRWN-00727369	GOOG-BRWN-00847948
GOOG-BRWN-00526782	GOOG-BRWN-00742713	GOOG-BRWN-00847949
GOOG-BRWN-00529122	GOOG-BRWN-00817563	GOOG-BRWN-00848368
GOOG-BRWN-00535100	GOOG-BRWN-00819269	GOOG-CABR-00000015
GOOG-BRWN-00536949	GOOG-BRWN-00819272	GOOG-CABR-00021811
GOOG-BRWN-00549861	GOOG-BRWN-00819441	GOOG-CABR-00051478
GOOG-BRWN-00554317	GOOG-BRWN-00819830	GOOG-CABR-00056264
GOOG-BRWN-00555073	GOOG-BRWN-00820370	GOOG-CABR-00057420
GOOG-BRWN-00555100	GOOG-BRWN-00826130	GOOG-CABR-00057779
GOOG-BRWN-00555223	GOOG-BRWN-00826401	GOOG-CABR-00057895
GOOG-BRWN-00561172	GOOG-BRWN-00826529	GOOG-CABR-00057918
GOOG-BRWN-00562313	GOOG-BRWN-00826530	GOOG-CABR-00058557
GOOG-BRWN-00571757	GOOG-BRWN-00826531	GOOG-CABR-00058751
GOOG-BRWN-00572220	GOOG-BRWN-00826532	GOOG-CABR-00058926
GOOG-BRWN-00601937	GOOG-BRWN-00826534	GOOG-CABR-00059431
GOOG-BRWN-00605636	GOOG-BRWN-00826535	GOOG-CABR-00059481
GOOG-BRWN-00613409	GOOG-BRWN-00826536	GOOG-CABR-00059774
GOOG-BRWN-00613801	GOOG-BRWN-00826537	GOOG-CABR-00059864
GOOG-BRWN-00613802	GOOG-BRWN-00826550	GOOG-CABR-00060364
GOOG-BRWN-00615433	GOOG-BRWN-00840745	GOOG-CABR-00063770
GOOG-BRWN-00630517	GOOG-BRWN-00844093	GOOG-CABR-00073873
GOOG-BRWN-00650016	GOOG-BRWN-00845312	GOOG-CABR-00073875
GOOG-BRWN-00663644	GOOG-BRWN-00845569	GOOG-CABR-00073878
GOOG-BRWN-00680546	GOOG-BRWN-00845585	GOOG-CABR-00073880
GOOG-BRWN-00697719	GOOG-BRWN-00845673	GOOG-CABR-00073922
GOOG-BRWN-00699213	GOOG-BRWN-00845676	GOOG-CABR-00086797
GOOG-BRWN-00699534	GOOG-BRWN-00846508	GOOG-CABR-00086881

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

GOOG-CABR-00101695	GOOG-CABR-00801323	GOOG-CABR-03841628
GOOG-CABR-00140102	GOOG-CABR-00832014	GOOG-CABR-03849022
GOOG-CABR-00148254	GOOG-CABR-00891629	GOOG-CABR-03907818
GOOG-CABR-00228748	GOOG-CABR-00892264	GOOG-CABR-03921989
GOOG-CABR-00358713	GOOG-CABR-00892455	GOOG-CABR-03922755
GOOG-CABR-00374449	GOOG-CABR-00901787	GOOG-CABR-03923580
GOOG-CABR-00377968	GOOG-CABR-00903677	GOOG-CABR-03923893
GOOG-CABR-00381312	GOOG-CABR-03611484	GOOG-CABR-03958924
GOOG-CABR-00391277	GOOG-CABR-03618845	GOOG-CABR-03959283
GOOG-CABR-00392675	GOOG-CABR-03622146	GOOG-CABR-03983354
GOOG-CABR-00399988	GOOG-CABR-03631556	GOOG-CABR-03983707
GOOG-CABR-00411167	GOOG-CABR-03646925	GOOG-CABR-03987880
GOOG-CABR-00421851	GOOG-CABR-03652549	GOOG-CABR-04005165
GOOG-CABR-00422906	GOOG-CABR-03652751	GOOG-CABR-04006287
GOOG-CABR-00427432	GOOG-CABR-03653330	GOOG-CABR-04007375
GOOG-CABR-00429070	GOOG-CABR-03662975	GOOG-CABR-04011132
GOOG-CABR-00430076	GOOG-CABR-03665840	GOOG-CABR-04073287
GOOG-CABR-00430662	GOOG-CABR-03667366	GOOG-CABR-04076570
GOOG-CABR-00487012	GOOG-CABR-03669893	GOOG-CABR-04077431
GOOG-CABR-00489377	GOOG-CABR-03670580	GOOG-CABR-04081967
GOOG-CABR-00501220	GOOG-CABR-03710192	GOOG-CABR-04082091
GOOG-CABR-00543418	GOOG-CABR-03710213	GOOG-CABR-04118321
GOOG-CABR-00543864	GOOG-CABR-03716577	GOOG-CABR-04122554
GOOG-CABR-00545997	GOOG-CABR-03717199	GOOG-CABR-04126517
GOOG-CABR-00547295	GOOG-CABR-03718983	GOOG-CABR-04134430
GOOG-CABR-00585454	GOOG-CABR-03738741	GOOG-CABR-04141174
GOOG-CABR-00710970	GOOG-CABR-03742530	GOOG-CABR-04141705
GOOG-CABR-00799341	GOOG-CABR-03751927	GOOG-CABR-04147060
GOOG-CABR-00800511	GOOG-CABR-03827263	GOOG-CABR-04177585

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

GOOG-CABR-04206179	GOOG-CABR-04635379	GOOG-CABR-04772394
GOOG-CABR-04207454	GOOG-CABR-04635593	GOOG-CABR-04773853
GOOG-CABR-04215325	GOOG-CABR-04639899	GOOG-CABR-04782100
GOOG-CABR-04240756	GOOG-CABR-04641725	GOOG-CABR-04782308
GOOG-CABR-04261880	GOOG-CABR-04648434	GOOG-CABR-04786706
GOOG-CABR-04293141	GOOG-CABR-04678169	GOOG-CABR-04795991
GOOG-CABR-04294727	GOOG-CABR-04687283	GOOG-CABR-04801490
GOOG-CABR-04307614	GOOG-CABR-04691939	GOOG-CABR-04804158
GOOG-CABR-04309395	GOOG-CABR-04692144	GOOG-CABR-04819485
GOOG-CABR-04324934	GOOG-CABR-04695140	GOOG-CABR-04820567
GOOG-CABR-04333674	GOOG-CABR-04695672	GOOG-CABR-04821454
GOOG-CABR-04408590	GOOG-CABR-04696292	GOOG-CABR-04828303
GOOG-CABR-04410117	GOOG-CABR-04697357	GOOG-CABR-04899841
GOOG-CABR-04417613	GOOG-CABR-04705124	GOOG-CABR-04959606
GOOG-CABR-04419756	GOOG-CABR-04706890	GOOG-CABR-04970427
GOOG-CABR-04423260	GOOG-CABR-04715843	GOOG-CABR-04986312
GOOG-CABR-04466637	GOOG-CABR-04716372	GOOG-CABR-05136994
GOOG-CABR-04470006	GOOG-CABR-04717190	GOOG-CABR-05145417
GOOG-CABR-04473461	GOOG-CABR-04720562	GOOG-CABR-05149122
GOOG-CABR-04477378	GOOG-CABR-04721001	GOOG-CABR-05163088
GOOG-CABR-04489649	GOOG-CABR-04722666	GOOG-CABR-05171191
GOOG-CABR-04508088	GOOG-CABR-04724084	GOOG-CABR-05173779
GOOG-CABR-04511160	GOOG-CABR-04732430	GOOG-CABR-05248101
GOOG-CABR-04606185	GOOG-CABR-04737403	GOOG-CABR-05269357.R
GOOG-CABR-04609856	GOOG-CABR-04743125	GOOG-CABR-05270014
GOOG-CABR-04616196	GOOG-CABR-04750983	GOOG-CABR-05280050
GOOG-CABR-04618590	GOOG-CABR-04761793	GOOG-CABR-05280756
GOOG-CABR-04624189	GOOG-CABR-04763333	GOOG-CABR-05280966
GOOG-CABR-04626237	GOOG-CABR-04763527	GOOG-CABR-05305495

CONFIDENTIAL - SUBJECT TO PROTECTIVE ORDER

GOOG-CABR-05362596
GOOG-CABR-05404845
GOOG-CABR-05405676
GOOG-CABR-05435664
GOOG-CABR-05454633
GOOG-CABR-05455683
GOOG-CABR-05461707
GOOG-CABR-05466323
GOOG-CABR-05468204
GOOG-CABR-05673102
GOOG-CABR-05737898
GOOG-CABR-05741188
GOOG-CABR-05756666
GOOG-CABR-05860125
GOOG-CABR-05864545
GOOG-CABR-05876612
GOOG-CABR-05876730
GOOG-CABR-05876763
GOOG-CABR-05876831
GOOG-CABR-05876933

Source Files:

- dbl-row-22ce947e70d000b1
- dbl-row-229a0a196fd000dc
- Consolidated 2022-04-08 3rd Round Search and Identifiers v2
- IP + UA Analysis (Source Material: GOOG-BRWN-00847947-948) (CONFIDENTIAL)